Replacing Missing Values in the Standard MISR Radiometric Camera-by-Camera Cloud Mask (RCCM) Data Product

Michel M. Verstraete[1], Linda A. Hunt[2], Hugo De Lemos[1], and Larry Di Girolamo[3]

[1]Global Change Institute (GCI), University of the Witwatersrand, Braamfontein, Republic of South Africa.
[2]Science Systems and Applications, Inc. (SSAI), Hampton, VA 23666-5845, USA.
[3]Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

Correspondence: Michel M. Verstraete (Michel.Verstraete@wits.ac.za or MMVerstraete@gmail.com)

Abstract. The Multi-angle Imaging SpectroRadiometer (MISR) is one of the five instruments hosted on-board the NASA Terra platform, launched on 18 December 1999. This instrument has been operational since 24 February 2000 and is still acquiring Earth Observation data as of this writing. The primary missions of MISR are to document the state and properties of the atmosphere, and in particular the clouds and aerosols it contains, as well as the planetary surface, on the basis of 36 data channels gathered by each of its nine cameras (pointing in different directions along the orbital track) in four spectral bands (blue, green, red and near-infrared). The Radiometric Camera-by-Camera Cloud Mask (RCCM) is derived from the calibrated measurements at the nominal top of the atmosphere, and is provided separately for each of the nine cameras. This RCCM data product is permanently archived at the NASA Atmospheric Science Data Center (ASDC) in Langley, VA, USA and is openly accessible (Diner et al. (1999b) and `https://doi.org/10.5067/Terra/MISR/MIRCCM_L2.004`). For various technical reasons described in this paper, this RCCM product exhibits missing data, even though an estimate of the clear or cloudy status of the environment at each individual observed location can be deduced from the available measurements. The aims of this paper are (1) to describe how to replace most missing values by estimates and (2) to briefly describe the software to process MISR RCCM data products, which is openly available to the community from the GitHub web site `https://github.com/mmverstraete` or `https://doi.org/10.5281/zenodo.3240018`. Limited amounts of updated MISR RCCM data products are also archived in South Africa and can be made available upon request.

1 Introduction

The Multi-angle Imaging SpectroRadiometer (MISR) is a unique instrument that features nine "pushbroom" imaging cameras, each equipped to acquire data in four spectral bands (blue, green, red and near-infrared), thereby offering 36 data channels to describe the properties of the Earth's environment (see Figure 1). These nine cameras, labeled DF, CF, BF, AF, AN, AA, BA, CA, and DA, are oriented at various bore angles along the orbital track of the Terra platform to acquire measurements at observers' zenith angles ranging from 70.3° forward to 70.6° aft. This instrument and its numerous applications have been amply
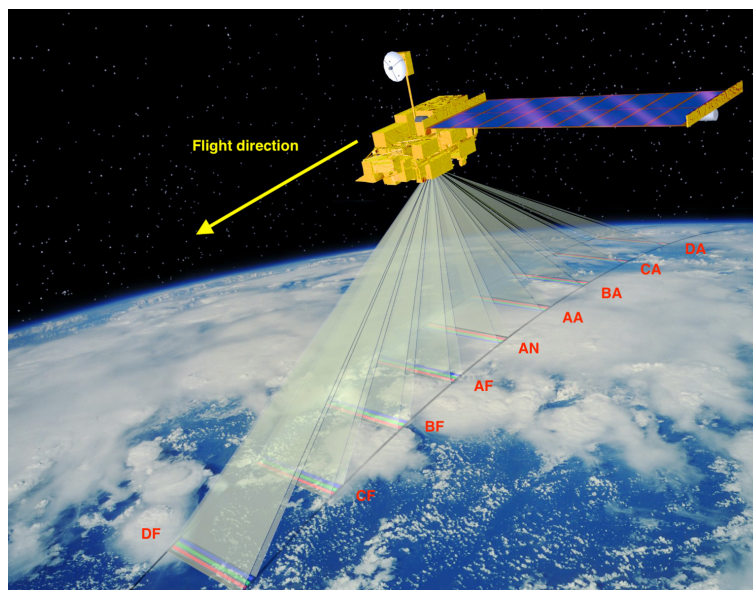
Figure 1. Artist rendition of MISR's observational protocol, with 9 cameras pointing in different directions along-track, each featuring 4 spectral bands. Image courtesy of Shigeru Suzuki and Eric M. De Jong, Solar System Visualization Project. Source: JPL image P-49081.

described in the literature (Diner et al., 1989, 1998, 1999a, 2005); a complete list of all publications about the MISR instrument and data products is available at `https://www-misr.jpl.nasa.gov/publications/peerReviewed/`.

The spectral detectors and the optical lens systems of the various MISR cameras are designed to yield an effective spatial resolution across track of 275 m in the eight off-nadir pointing cameras. For cost-related reasons, the nadir pointing camera

5 AN is identical to the next two off-nadir pointing cameras AF and AA; which results in a slightly finer spatial resolution of 250 m. However, those data are re-sampled to 275 m as part of the Level 1 processing. Furthermore, to reduce the amount of data that need to be transmitted to the ground segment, raw data are averaged to a spatial resolution of 1100 m in the non-red data channels of the off-nadir pointing cameras. Hence, the four spectral bands of the nadir pointing camera, and the red spectral band in the eight off-nadir cameras, are available at the native spatial resolution of 275 m, while all other data channels are

10 only available at the reduced spatial resolution, in the default Global Mode of operation. MISR can occasionally be operated in Local Mode, in which case all 36 data channels are available at the native spatial resolution of 275 m for a limited area along track, but this type of acquisition will not be considered further in this paper as the cloud mask discussed in this paper is only derived in Global Mode.

The orbit of the Terra platform is carefully managed to maintain its altitude as well as the Equatorial Crossing Time (ECT)

15 on its way from the North Pole to the South Pole. The original specification called for no more than 15 minutes deviation from the nominal 10:30 AM local time, but was tightened to within one minute after two years of operations. This platform completes sequences of 233 orbits in 16 days, after which it returns to observe the planet from the same relative position. In the MISR context, every orbit thus falls within one of these 233 fixed patterns, called "Paths". Each orbit is numbered sequentially,

starting with 1 at launch. The MISR instrument collects data only between the northern and southern terminators, while over-flying the sunlit side of the Earth, and the data acquired during that time are referred to as an "Orbit", even though it actually comprises only about half of the orbit. MISR became operational on 24 February 2000, while performing Orbit 995, and has already completed over 100,000 Orbits; it is still operational as of this writing, and therefore offers a unique opportunity to

5 study environmental issues over a continuous range of 19+ years.

To facilitate the analysis of MISR data, the maximum possible daylit extent in each Orbit is segmented in 180 "Blocks" that have a fixed geographical location (as long as the orbit is properly maintained). Each Block is a rectangular area of 563.2 km across-track by 140.8 km along-track, which is large enough to contain the swath of the instrument. Successive Blocks of a given Orbit are staggered with respect to each other to account for the inclined orbit of the platform with respect to the

10 Equatorial plane. For any given Orbit, only about 142 Blocks contain actual data (Bull et al., 2011, p. 306).

Lastly, at Level 1B2, each and every individual MISR measurement (i.e., pixel value) is coded as a 16-bit unsigned integer, where the 14 most significant bits carry the scaled radiance information, and the last 2 bits contain the Radiometric Data Quality Indicator (RDQI), an indicator of the quality of that particular measurement. This RDQI can take on the following values (Bull et al., 2011, p. 190):

15   – Binary `00`, integer `0`: the pixel value is "Within specifications".

    – Binary `01`, integer `1`: the pixel value is deemed of "Reduced accuracy".

    – Binary `10`, integer `2`: the pixel value is deemed "Not usable for science".

    – Binary `11`, integer `3`: the pixel value is "Unusable for any purpose".

In this paper and the associated software, those values are labeled "good", "fair", "poor" and "bad", respectively.

20 **2 Motivation**

The standard MISR Radiometric Camera-by-Camera Cloud Mask (RCCM) data product (Diner et al. (1999b) and `https://doi.org/10.5067/Terra/MISR/MIRCCM_L2.004`) assigns an 8-bit unsigned integer value to each pixel within each Block. Different algorithms are applied over oceanic and terrestrial areas to determine the state of cloudiness at each geographical location, at spatial resolution of 1.1 km. The RCCM is generated for each of the nine cameras using two different

25 spectral bands, namely the red and near-infrared bands. All results discussed in this paper refer to the RCCM product Version F04_0025, which is the most current version available at the time of writing.

Over oceans, one of the tests used to generate the RCCM relies on the bidirectional reflectance factor (BRF) in the near-infrared band at 1.1 km resolution, while the other examines the standard deviation of the 4×4 array of the 275 m red band BRF within a 1.1 km squared area. The near-infrared BRF and the red band standard deviation results for each pixel are each

30 tested against three thresholds to classify a pixel as high-confidence cloudy, low-confidence cloudy, low-confidence clear, or high-confidence clear. The two tests may return different results, and the final cloud mask is determined from the logical combination of the results of the two tests.

Over land, the standard deviation test remains the same as over ocean, but the near-infrared BRF test is replaced with a vegetation index constructed from a combination of red and near-infrared BRFs. The logic for combining the two tests is different over land than it is over ocean, with a lower reliance on the standard deviation test over land. These algorithms are described in detail in the Level 1 Cloud Detection Algorithm Theoretical Basis Document (ATBD) (Diner et al., 1999b),

5 which can be downloaded from `https://eospso.gsfc.nasa.gov/atbd-category/45`. Updates to the RCCM are provided in (Zhao and Di Girolamo, 2004) and (Yang et al., 2007). Validation is described within the GEWEX Cloud Assessment Report (Stubenrauch et al., 2012), which also shows that the RCCM compares favorably to many other satellite products when used to compute cloud fraction over snow- and ice-free surfaces.

The RCCM values carry the following meanings (Bull et al., 2011, p. 85):

10   – `0B`: No retrieval.

    – `1B`: Cloud with high confidence.

    – `2B`: Cloud with low confidence.

    – `3B`: Clear with low confidence.

    – `4B`: Clear with high confidence.

15   – `255B`: Fill value.

where the `B` notation indicates that those values are unsigned 8-bit (or single byte) integer values.

This data product typically contains missing (no retrieval) data in the following four cases:

1. As part of the Level 1 processing, measurements are re-projected to a Path-specific Space Oblique Mercator (SOM) map projection, which minimizes distortions and scale errors (see, in particular, Appendix A of Bull et al., 2011,

20   p. 299). This re-projection is implemented both on the World Geodetic System 1984 (WGS84) reference ellipsoid (for oceanic applications and is referred to as Ellipsoid projected) and on a representation of the actual topography of the continents (referred to as Terrain projected). Over ocean, Terrain and Ellipsoid projected radiances are equivalent, so Terrain projected radiances are not reported over ocean. Within any camera, but most notably for those observing the Earth at larger zenith angles, some ground areas may be obscured by the local terrain in the Terrain projected radiance

25   (e.g., the surface area may lie behind hills or mountains, and may be unobservable from the point of view of those cameras—hence, no Terrain projected radiance are reported for those cameras).

2. The data files containing RCCM data cover geographical areas that are wider than the effective swath width of the MISR instrument. Hence, in any particular Block of data, the western and eastern sides of the rectangular Block are filled with a particular code indicating where no observations are available: these are referred as the *swath edge* pixels.

30 3. Occasionally, the on-board computer gets overwhelmed by the amount of data to be ingested. This may occur for various reasons, but happens most frequently when switching between Global and Local Mode, or conversely. Whenever this

occurs, the computer quickly resets itself and restarts operations. These events are recorded as part of the initial raw data processing, and result in one or more *missing lines* of measurements. After re-projection onto the SOM map, these dropped lines appear as curved lines across the entire Block in the data.

4. The software processor that interprets MISR data and generates the RCCM data product currently considers only L1B2 radiance data with an RDQI of `0B`, and therefore ignores and dismisses all measurements considered of lower quality.

In the first two cases, no observations are made so these are truly missing values for which no estimate can be provided. In the third case, only missing values in the red or near-infrared spectral bands result in missing RCCM values, However, these missing lines affect different geographical locations in different spectral bands. This is because the red and NIR detectors lie next to each other on the camera focal planes, so that, at any given time, they actually observe locations separated by a few km on the ground. This shift in geolocation is taken care of during the L1B1 processing stage. The interesting outcome is that if an event results in the simultaneous dropping of a data line in both spectral bands, different locations are affected on the Earth's surface. In the fourth case, any radiance value in the red or NIR spectral band with an RDQI of `1B` or higher would cause the RCCM to be missing over land, even though the cloudiness status of the area could be derived on the basis of available data.

Missing lines in L1B2 data have different implications for RCCM over water bodies and land masses. Over oceans, since the two tests use different spectral bands, one of them is often capable of estimating the cloudiness status of each pixel, because the geographical locations affected by missing lines varies with each spectral band. In the case of land, the vegetation index test uses the red and the NIR band, so if either of those bands is unavailable, the vegetation test fails automatically. In this situation, the logic for combining the two tests over land will report Cloud with High Confidence only if the red band is available, contains 9 or more of the 16 275m samples for computing the standard deviation of the red-band reflectance, and passes the Cloud with High Confidence threshold; otherwise it is flagged as "No Retrieval". Hence, there are many more cases of RCCM missing values over land. These various cases are exhibited in the following three figures, for the south-eastern coast of South Africa and the Indian Ocean.
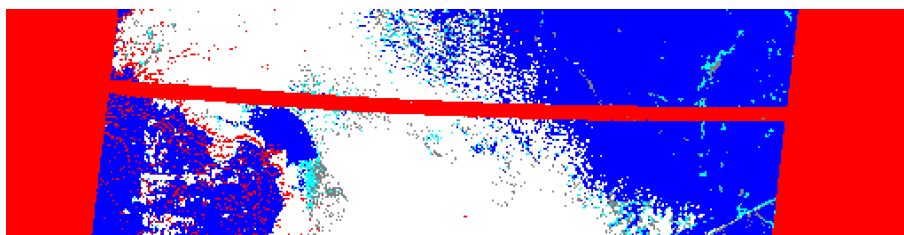
Figure 2. Map of the MISR RCCM product for Path 168, Orbit 68283, Block 110 and camera CA. All RCCM products are generated at the spatial resolution of 1100 m and provided as arrays of 512 by 128 pixels. These maps have been enlarged (4x in each direction) by duplication for viewing convenience and to facilitate comparisons with other maps. The total size of the Block area is 563.2 km across-track by 140.8 km along-track, while parallelogram-shaped ground area inside the Block is about 380 km across-track. Color coding: red: no retrieval (RCCM = `0B`) or fill value (RCCM = `255B`); white: cloud with high confidence (RCCM = `1B`); gray: cloud with low confidence (RCCM = `2B`); aqua: clear with low confidence (RCCM = `3B`); and blue: clear with high confidence (RCCM = `4B`).

Figure 2 shows a map of a particular Block of standard MISR RCCM data, where red pixels are deemed non-retrieved or "missing". The scattered isolated red dots within the mapped area correspond to locations that cannot be acquired by that particular camera (the observation zenith angle for camera CA is 60.2°) due to obscuration by steep topography in the Drakensberg mountain range. The red areas on the western and eastern ends of the Block correspond to the swath edge areas:

5 they fall outside the radiance swath of the MISR instrument for that camera and contain only fill values. The wide red curved line that runs across the entire area is an example of the effect of missing values in the red or near-infrared spectral bands of the MISR L1B2 data product (in this case, entirely over land).

Figure 3 shows a different example, in this case from another combination of Orbit and Block numbers. There are fewer scattered red dots because this map was derived for a less inclined camera (BF, with an observation zenith angle of 45.7°)

10 observing a somewhat smoother terrain. There are again a couple of red curved lines running across the entire area, due to the missing red and near-infrared L1B2 radiance data. But there is also a linear feature aligned along-track near the western edge of the Block: in that case, the L1B2 radiance data are not actually missing, but associated with an RDQI of 1B.
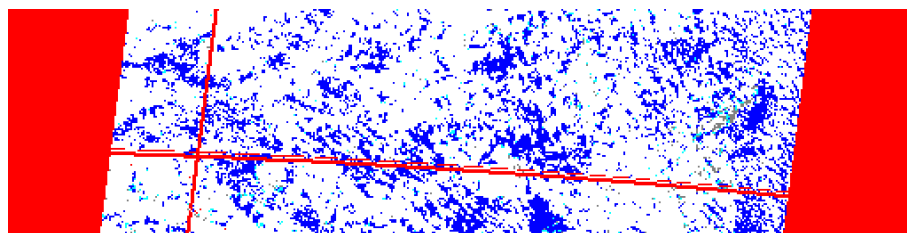
Figure 3. Map of the MISR RCCM for Path 168, Orbit 1412, Block 108 and camera BF. The linear dimensions of the mapped area and the color coding are identical to those indicated in Figure 2.
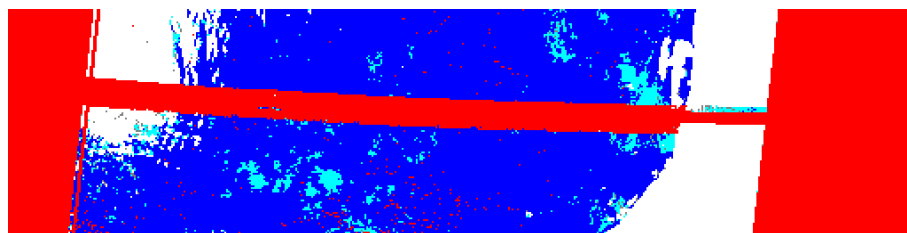
Figure 4. Map of the MISR RCCM for Path 168, Orbit 98340, Block 113 and camera BA. Note the variable thickness of the missing data band when crossing the shoreline between the continent and the Indian Ocean (See text for details). The linear dimensions of the mapped area and the color coding are identical to those indicated in Figure 2.

Figure 4 exhibits a case where the numbers of missing lines in the red and NIR spectral bands of the L1B2 data are so large that they overlap on the Earth's surface. It can be seen that the thickness of the band of missing data in the RCCM product is

15 larger over land than over the ocean. This is because one of the tests could work over water when only one spectral band was

available, while neither tests were operational over land. Note also the presence of another straight line of missing values along the western edge of the Block, due to L1B2 radiance values associated with an RDQI of 1B.
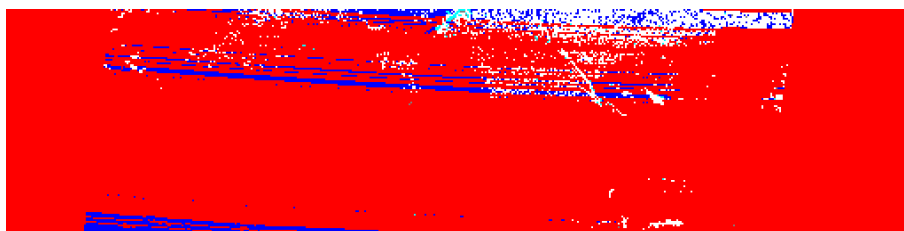
Figure 5. Map of the MISR RCCM for Path 168, Orbit 2111, Block 110 and camera BA. The linear dimensions of the mapped area and the color coding are identical to those indicated in Figure 2.

Figure 5 exhibits an extreme case of missing RCCM data: Here, most of the missing cloud mask values are due to L1B2 radiance data associated with an RDQI of 1B or higher, as can be seen from Figure 6, showing the RDQI values associated with the L1B2 radiance data in the red spectral band, one of the two used to determine the RCCM.
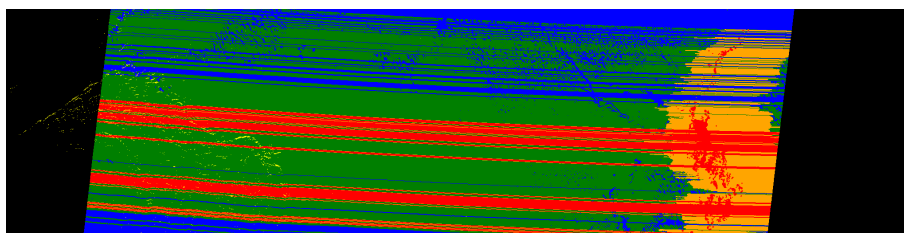
Figure 6. Map of the MISR L1B2 radiance quality data for Path 168, Orbit 2111, Block 110 and camera BA, in the red spectral band. This particular data channel is always available at the spatial resolution of 275 m. The linear dimensions of the mapped area are identical to those indicated in Figure 2. Color coding: black: edge pixels; yellow: area obscured by the local topography; blue: good data (RDQI = 0), green: fair data (RDQI = 1), gold: poor data (RDQI = 2), red: bad or missing data (RDQI = 3).

To be fair, it is important to note that missing L1B2 radiance data affect only a small proportion of all globally available data over any significant amount of time. However, these dropped lines tend to occur more systematically in connection with the switching between Global and Local Mode, as mentioned earlier, so they very much affect the usability of the products at those locations. Figure 7 shows how many pixel values went missing in the 9 cameras of the RCCM product for Path 168 and Block 110, which includes the Skukuza flux tower, a site initially monitored in Local Mode early in the mission in support of the Safari-2000 field campaign (see the special issue of the Journal of Geophysical Research dedicated to that campaign, openly available from https://agupubs.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)2169-8996.SAF1), and systematically acquired since December 2009. A close inspection of the data reveals that straight lines of L1B2 radiance data with an RDQI value of 1B contaminated cameras DF and BF of just about every Orbit between the start of the mission and the end of March 2008. Between April 2008 and December 2009, four cameras were affected in this way: CF, BF, BA and DA. Very few missing

RCCM values occur during a short period of time (late December 2009 to early March 2010), even though this corresponds to the start of the systematic acquisition of Local Mode observations for that particular Path and Block. A highly variable number of missing values have been observed since March 2010.
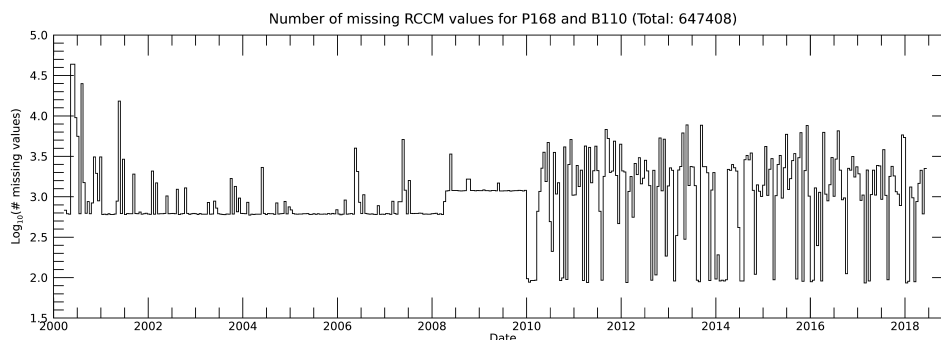
Figure 7. Plot of the number of missing RCCM values for Path 168 and Block 110, as a function of time (i.e., for all available Orbits) between the start of the mission and mid-2018. This is a semi-logarithmic plot, so each vertical step of one unit represents an order of magnitude more missing points (See text for details).

Based on inspecting these and many other cases, we conclude that (1) some of the missing RCCM data result from actually
5 missing L1B2 radiance data, in the form of dropped lines showing up as curved bands across the swath of the instrument, (2) missing RCCM data also arise in conjunction with the presence of L1B2 radiance data associated with an RDQI of 1B or higher, which has historically contaminated a large number of Orbits, especially between the start of the mission and December 2009, and (3) whenever measurements are missing, they may affect multiple cameras and spectral bands, but the geographical areas affected are different for each camera and band, due to the hardware configuration of the camera focal planes.

10 This suggests that analyzing the RCCM data in multiple cameras and spectral bands, or assuming some degree of spatial continuity in cloud fields, might prove useful in establishing the clear or cloudy status of a geographical location where the standard RCCM data product available from the NASA Atmospheric Science Data Center (ASDC) in Langley, VA, USA, is missing.

The purpose of this paper is to describe a set of simple algorithms and associated software codes (publicly available from the
15 GitHub web site https://github.com/mmverstraete or https://doi.org/10.5281/zenodo.3240018) that can be applied to that standard RCCM data product to generate an updated version of that product where most if not all of the missing values of type 3 and 4 above have been replaced by estimates of the clear or cloudy status, based on the values actually available in other cameras and spectral bands. In turn, a more spatially complete RCCM product may be useful to users of those data or derived products.

3 Processing

It is proposed to replace the missing values in the standard MISR RCCM data product by estimates of the clear or cloudy status of individual geographical locations in four successive steps, each implemented in a separate function written in the IDL language (IDL is a commercial software product distributed by Harris Geospatial Solutions, Inc.)

5   For each of the 9 MISR cameras, the following steps are implemented:

3.1 Reading data

The first step, implemented in IDL function `mk_rccm0.pro`, reads the standard MISR RCCM data product from the appropriate files for the specified Path, Orbit, Block and camera, downloaded from the NASA ASDC.

As indicated above, this original data set assigns a "No Retrieval" value to all RCCM pixels corresponding to MISR L1B2
10 radiance values associated with an RDQI other than `0B`, independently from the reason for this latter quality indicator: This standard product therefore does not distinguish between unobservable locations (e.g., swath edge pixels or obscured by terrain) and genuinely missing observations.

3.2 Relabeling data

The second step, implemented in IDL function `mk_rccm1.pro`, accesses the standard MISR L1B2 scaled radiance and RDQI
15 data products in the 4 spectral bands for the same Path, Orbit, Block and camera, in this case through pointers to data arrays that have been previously copied on the heap for efficiency reasons. The RCCM data array for the current camera is then updated as follows:

- If a pixel is deemed obscured (scaled radiance value with the RDQI attached of `65511`) in any one of the 4 spectral bands of the current camera, then that pixel value is set to `253B` in the RCCM data array.

20 - If a pixel is classified as belonging to the swath edges of the Block (scaled radiance value with the RDQI attached of `65515`) in any one of the 4 spectral bands of the current camera, then that pixel value is set to `254B` in the restored RCCM data array.

At this point, all remaining `0B` pixel values in the RCCM array for the current camera indicate missing values that might be replaced by estimates.

25 ### 3.3 Replacing missing values (same location, neighboring cameras)

In the third step, implemented in IDL function `mk_rccm2.pro`, the geographical locations of missing pixels in the current camera are retrieved. For each such missing RCCM value, the corresponding values at the very same geographical location (sample and line numbers within the Block) *in the preceding and following cameras*, in the natural sequence DF to DA mentioned earlier, are inspected. If these values are valid (not missing) and equal, i.e., if their values are in the range [`1B`,
30 `4B`] and identical in both cameras, then that same value is also assigned to the missing value in the current camera. This

assignment assumes that if that particular location was regularly observed by the previous and the following cameras, and both of these cameras assign the same RCCM code defined in Section 2, then that location would likely have been assigned that same value if the measurement had not been missing from the current camera.

This part of the algorithm calls for two comments:

5   – Clouds always occur at some altitude above the ground. Hence, when different MISR cameras point to a particular location on the digital elevation model used in the terrain projection (as explained in Section 2), they may actually observe different atmospheric volumes in altitude (parallax effect). This decision rule therefore assumes that the clear or cloudy areas observable by those neighboring cameras are actually continuous over the range of observation angles spanned by these three cameras. This rule is very effective and efficient in the case of large clear zones or consistent (overcast)

10   cloud decks, but may lead to erroneous settings in the particular case where the cloud field is very discontinuous, i.e., composed of clouds (or clear areas) small enough to affect one camera but not its immediate neighbors. This limitation is mitigated by the fact that *two* neighboring cameras are inspected: if there is any discrepancy between their RCCM values, the situation is skipped at this stage.

    – The two extreme cameras (DF and DA) do not have both a preceding or a following camera. In those cases, the two

15   cameras following the DF camera (CF and BF) and the two cameras preceding DA (BA and CA) are inspected instead, using the same approach.

This step only addresses cases where there is consistency between the neighboring cameras regarding the clear or cloudy status of the locations for which there is missing information in the current camera. The remaining cases are more subtle, as they arise when neighboring cameras provide divergent assessments of the situation.

20   **3.4   Replacing missing values (neighboring locations, same camera)**

In this last step, implemented in IDL function `mk_rccm3.pro`, a decision concerning the probable clear or cloudy status of the remaining missing pixels of type 3 and 4 is based on the values of neighboring pixels *within the current camera*. This final assignment is implemented in 4 stages.

3.4.1   Stage A

25   The RCCM field for the current camera is first scanned to locate the remaining missing values. For each one of them, a small sub-window of 3 by 3 pixels is defined, centered on the missing pixel, and the valid RCCM values within that sub-window are inspected. If there are at least 4 valid (non-missing) values within that sub-window, *and* if all of them exhibit identical RCCM values, then that same value is assigned to the missing value. This rule is applied repeatedly, until no more missing values can be replaced using that decision rule. Hence, each time the Block is scanned to locate missing pixels meeting those

30   requirements, values that have been reassigned previously can influence the assignment of neighboring missing RCCM values.

3.4.2 Stage B

If some missing values remain, then a new sub-window of 5 by 5 pixels is defined, again centered on each of the missing pixels. This time, a minimum of 12 valid values are required to make a decision, and they are allowed to differ from each other. The minimum (`min`), median (`med`) and maximum (`max`) RCCM values among those valid values are computed, and the following

5 decision rules are applied:

- If `min` = 1B, `max` = 2B, and `med` < 1.5, then the missing value is set to 1B.

- If `min` = 1B, `max` = 2B, and `med` ≥ 1.5, then the missing value is set to 2B.

- If `min` = 2B, `max` = 3B, and `med` < 2.5, then the missing value is set to 2B.

- If `min` = 2B, `max` = 3B, and `med` ≥ 2.5, then the missing value is set to 3B.

10 - If `min` = 3B, `max` = 4B, and `med` < 3.5, then the missing value is set to 3B.

- If `min` = 3B, `max` = 4B, and `med` ≥ 3.5, then the missing value is set to 4B.

- If `min` = 1B, `max` = 3B, and `med` < 1.5, then the missing value is set to 1B.

- If `min` = 1B, `max` = 3B, and 1.5 ≤ `med` < 2.5, then the missing value is set to 2B.

- If `min` = 1B, `max` = 3B, and `med` ≥ 2.5, then the missing value is set to 3B.

15 - If `min` = 2B, `max` = 4B, and `med` < 2.5, then the missing value is set to 2B.

- If `min` = 2B, `max` = 4B, and 2.5 ≤ `med` < 3.5, then the missing value is set to 3B.

- If `min` = 2B, `max` = 4B, and `med` ≥ 3.5, then the missing value is set to 4B.

- If `min` = 1B, `max` = 4B, and `med` < 1.5, then the missing value is set to 1B.

- If `min` = 1B, `max` = 4B, and 1.5 ≤ `med` < 2.5, then the missing value is set to 2B.

20 - If `min` = 1B, `max` = 4B, and 2.5 ≤ `med` < 3.5, then the missing value is set to 3B.

- If `min` = 1B, `max` = 4B, and `med` ≥ 3.5, then the missing value is set to 4B.

In effect, these rules are designed to assign an RCCM value consistent with the statistical distribution of values within the window under consideration, namely to assign the cloudiness level closest to the median of the neighboring values.

If the missing value is located near the northern or southern border of the Block, there may be fewer than the 9 or 25 values
25 theoretically available in the 3 by 3 or 5 by 5 sub-windows. In those cases, the sub-windows are truncated and only valid values in those smaller sub-windows are considered. As before, this rule is repeatedly applied until no more missing values can be replaced.

3.4.3 Stage C

If the updated RCCM data product still contains missing values, the same algorithm as in stage B is applied, but this time requiring only 10 valid values to make a decision. As before, this rule is repeatedly applied until no more missing values can be replaced.

3.4.4 Stage D

Lastly, if there are some missing values left after stage C, the same algorithm is again applied, this time using the smaller 3 by 3 sub-window and requiring only 3 valid values in order to decide the clear or cloudy status of the missing value. As before, this rule is repeatedly applied until no more missing values can be replaced.

If there are still missing values after this last stage, the clear or cloudy status of the corresponding pixels is deemed impossible to determine. This actually occurs in only a handful of cases, when a very small area, corresponding to a single pixel in the reduced resolution data (hence of the order of 1.1 km by 1.1 km) is buried in mountainous terrain with fewer than 3 valid pixels in its immediate neighborhood.

3.5 Main function

The orderly application of those functions and the generation of outcomes is managed by the IDL function `fix_rccm.pro`, while the proper dimensioning of the sub-windows and the decision tree described above are implemented in another function called `repl_box.pro`.

In summary, the RCCM array for each camera updated through this process takes on the following values:

- `0B`: No retrieval.
- `1B`: Cloud with high confidence.
- `2B`: Cloud with low confidence.
- `3B`: Clear with low confidence.
- `4B`: Clear with high confidence.
- `253B`: Obscured pixel.
- `254B`: Edge pixel.
- `255B`: Fill value.

4 Outcomes

The RCCM data arrays that have been updated in this way contain very few missing values, if any, and also correctly report on the location of obscured and edge pixels. This same function `fix_rccm.pro` also offers the opportunity to map the results

obtained at the end of each step. For instance, the final updated RCCM data product for the same conditions as shown earlier in Figure 2 is exhibited in Figure 8.
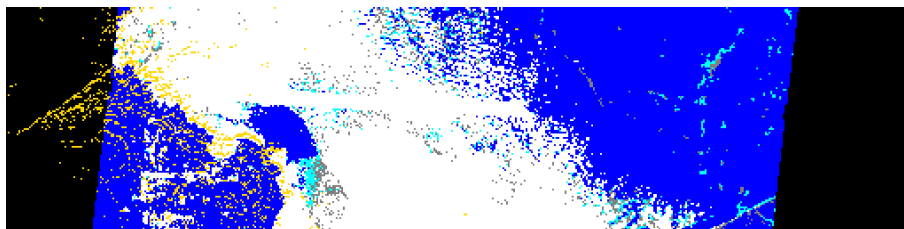
Figure 8. Map of the final updated (Stage D) MISR RCCM for Path 168, Orbit 68283, Block 110 and camera CA. Missing pixels have been replaced by estimates of the cloud or clear status of the observed areas, based on the cloudiness level of neighboring cameras and neighboring pixels within small sub-windows of the target camera, as described in the text. The remarks and the linear dimensions of the mapped area are identical to those indicated in Figure 2. Color coding: red: no retrieval or fill value; white: cloud with high confidence; gray: cloud with low confidence; aqua: clear with low confidence; blue: clear with high confidence; gold: pixels obscured by topography; and black: pixels in the edges of the instrument swath.

It can be seen that the algorithm was able to correctly discriminate between the largely cloudy area in the middle of the scene and the mostly clear areas near both edges. It has also quite conservatively assigned most of the pixels within the main cloudy area to "cloudy with high confidence": this is, in part, the result of more inclined cameras observing more continuous cloud fields, and often constitutes the more conservative assessment (Zhao and Di Girolamo, 2004).

Figure 9 similarly exhibits the outcome of the process for the case shown earlier in Figure 3. In this case, the relative narrowness of the missing lines permitted the algorithm to restore most or all of the fine structure of the broken cloud field, without generating any noticeable artifact.
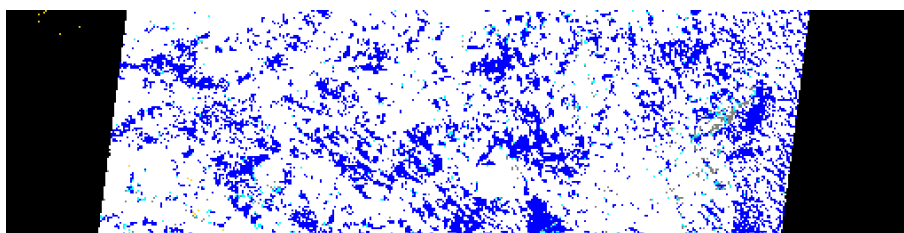
Figure 9. Map of the final updated (Stage D) MISR RCCM for Path 168, Orbit 1412, Block 108 and camera BF. Missing pixels have been replaced by estimates of the cloud or clear status of the observed areas, based on the cloudiness level of neighboring cameras and neighboring pixels within small sub-windows of the target camera, as described in the text. The remarks and the linear dimensions of the mapped area are identical to those indicated in Figure 2, while color coding is identical to that indicated in Figure 8.

The performance of the process to replace missing RCCM values by estimates is really highlighted in the extreme case shown earlier in Figure 4, which has been restored as shown in the middle frame of Figure 10. To understand how that is

possible, the RCCM values for the previous and following cameras (AA and CA, respectively), are also exhibited. It can be seen that neither of those cameras is affected by missing values, so the algorithm described in subsection 3.3 is capable of restoring RCCM values for large consistently clear and cloudy areas. In fact, the outcome of that process (i.e., before applying the decision rules defined in subsection 3.4), can be seen in Figure 11.
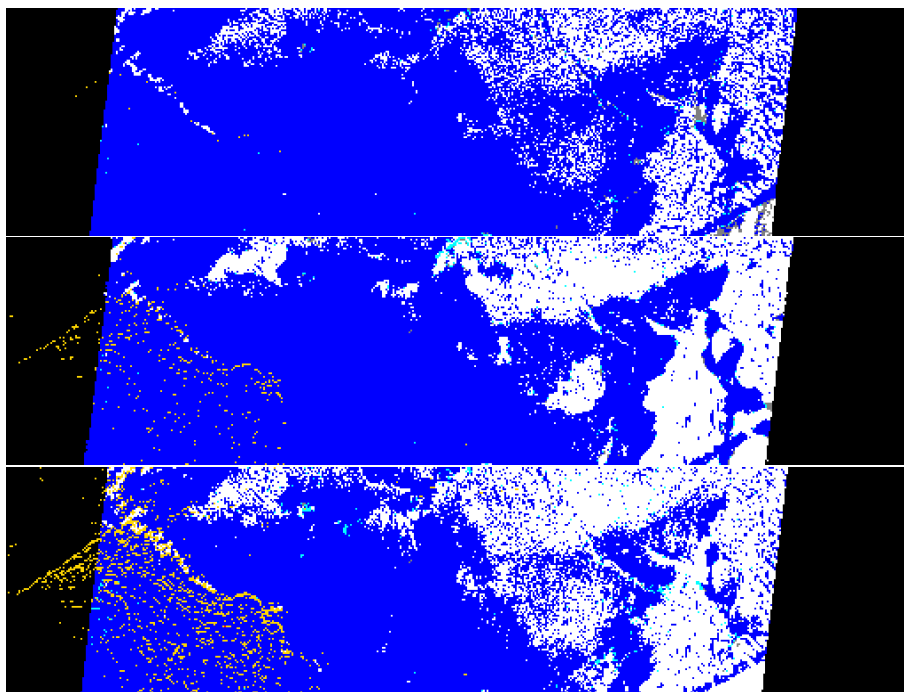
Figure 10. Maps of the MISR RCCM for Path 168, Orbit 2111, Block 110 and Cameras AA (top frame), BA (middle frame) and CA (bottom frame). Restoring the clear and cloudy areas in camera BA (central frame), despite the very large number of missing values (as shown in Figure 4), is greatly helped by the absence of missing RCCM data in the AA (top) and CA (bottom) cameras. Note also the increased cloud cover, as well as the more prominent impact of the local topography, when the zenith angle of observation increases. Linear dimensions are as described in the legend of Figure 2, while color coding is identical to that indicated in Figure 8.

5 Table 1 reports on the efficiency of the process to replace missing RCCM values in the various cases presented earlier in this paper. It can be seen that essentially all missing values are successfully replaced by one of the four possible discrete estimates of cloudiness. The same rate of success has been found in all other cases inspected.

5 Conclusions

The RCCM data product provides information on the spatial distribution of clouds in each of the 9 cameras of the MISR
10 instrument. This data product contains missing values, which can be attributed to four different causes. Terrain obscured and swath edge pixels are actually not observable but are now properly identified in the corrected data product.
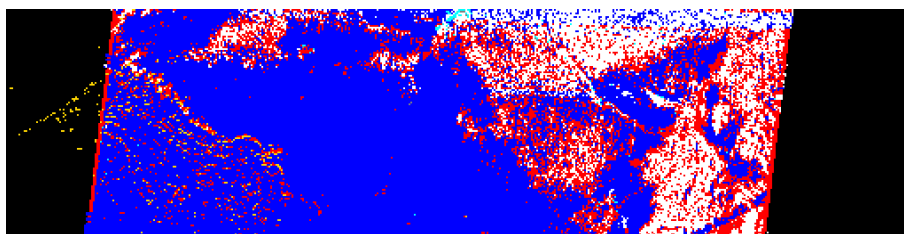
Figure 11. Map of the MISR RCCM for Path 168, Orbit 2111, Block 110 and camera BA. This map exhibits the intermediary product obtained at the end of the processing step involving values at the same location in the neighboring cameras (Subsection 3.3), and before investigating the values of immediate neighbors in the same camera (Subsection 3.4). Note also the presence of a straight line of missing values near the western edge of the Block, which is correctly replaced in the final corrected RCCM product shown in the middle frame of Figure 10. Linear dimensions are as described in the legend of Figure 2, while color coding is identical to that indicated in Figure 8.

Table 1. Number of missing RCCM values remaining at the end of steps 1, 2 and 3 for the cases exhibited elsewhere in this paper.

Orbit	Block	n1	n2	n3	Success rate
01412	108	4947	2156	0	100.00%
02111	110	43606	9555	1	99.99%
68283	110	5337	1142	1	99.98%
98340	113	7378	2184	0	100.00%

Missing values in the L1B2 radiance data product constitutes the third cause for missing RCCM values, as the latter is derived from an analysis of the former. This tends to happen when the MISR instrument is switched from Global to Local Mode and conversely. Such changes in acquisition Mode affect only a small number of Paths and Orbits throughout the mission, but since some sites (such as the Skukuza site in the Kruger National Park) are systematically acquired at this full spatial resolution for
5 scientific or calibration purposes, the areas surrounding those sites are then consistently contaminated by missing lines of data.

The systematic presence of L1B2 radiance data with an RDQI value higher than `0B`, at least during the first 9 years of the MISR mission, constitutes the fourth process that leads to missing RCCM values. This condition may occur when the data acquired by a camera may be affected by sun glint, for instance, or whenever the result of resampling the raw data at Level 1 calls for a warning (Veljko Jovanovic, MISR Team at JPL, personal communication on 26 March 2019). More details can
10 be found in the Algorithm Theoretical Basis document (Bruegge et al., 1999). In the experience of the authors, measurements associated with an RDQI of `1B` are perfectly suitable for research applications, and their systematic exclusion in the generation of the RCCM is unwarranted. The procedures described above and the associated computer codes available from GitHub will effectively deal with those missing RCCM values, though it may be more consistent—in the context of a future reprocessing exercise—to allow the original RCCM algorithms to accept data with an RDQI of `1B` in the first place.
15 The first two processes are entirely legitimate and missing data cannot be replaced by any reasonable value, however the algorithmic processing described in this paper results in the proper identification of those locations in the updated RCCM data

product. The last two processes can be handled and replacement values can, in most cases, be proposed to fill in the missing RCCM values. The probable clear or cloudy state of those locations is derived from an analysis of valid (i.e., non-missing) RCCM values at the same location in the two neighboring cameras, and/or at neighboring locations within the same camera.

5   Specific examples have been exhibited, including an extreme case where about 87% of the expected values in a RCCM data product are missing from a particular camera. In all cases, the RCCM values have been restored to visually reasonable and statistically conservative values. The only cases where the algorithms described in this paper cannot suggest a replacement value concern areas of size 1.1 km by 1.1 km (i.e., corresponding to a single pixel in the mapped product), isolated within mountainous areas, and with fewer than 3 neighboring pixels with valid values within the 3 by 3 or 5 by 5 sub-windows centered on the missing value itself.

10   These algorithms have been applied and tested extensively. As indicated earlier, all MISR L1B2 and RCCM data are freely and openly available from the NASA ASDC, and the IDL source codes used to generate the results described above are available under the MIT licence from the GitHub web site mentioned both above and below.

6   Code availability

This paper describes Version 2.15 of the IDL functions available in the `MISR_RCCM` repository of the GitHub web site `https://github.com/mmverstraete` or `https://doi.org/10.5281/zenodo.3240018`. These functions are part of a larger library of software routines to process MISR data and MISR-HR products. Please note that those tools are under active development and may be updated in time.

Each function contained in that repository contains abundant in-line documentation to describe every step in the processing. An HTML file with properly formatted documentation for each function in a standardized setup is also available from the same repository. These functions may also depend on other, more basic, utility functions defined in other repositories of the same GitHub web site. It is thus recommended to download and install the contents of all repositories to have access to these and other functionalities.

7   Data availability

All MISR data, including the RCCM and the L1B2 radiance data products mentioned in this paper, were obtained from the NASA Langley Research Center Atmospheric Science Data Center. These data sets are openly available from the ASDC web site at: `https://eosweb.larc.nasa.gov/project/misr/misr_table`. The references for those data sets are Diner et al. (1999b) (`https://doi.org/10.5067/Terra/MISR/MIRCCM_L2.004`) for RCCM and Jovanovic et al. (1999) (`https://doi.org/10.5067/Terra/MISR/MI1B2T_L1.003`) for L1B2.

Sample availability. Samples of updated MISR RCCM data products with missing data replaced by estimated values can be obtained by sending an email request to: `misrhrwits@gmail.com`.

Author contributions. MMV: Co-development of the algorithm, software testing, senior author in charge of the initial and final drafting of the manuscript; LAH: co-development of the algorithm, manuscript updating; HDL: MISR data download, local processing, archiving of results for South Africa.; LDG: description of standard RCCM algorithms, manuscript checking and updating.

Competing interests. The co-authors of this manuscript declare that they have no competing or conflicting interests regarding the content of
5 this paper.

Disclaimer. The IDL source code software mentioned above (obtainable from the GitHub web site) is made available under the MIT license, which states, in part, that "The software is provided *as is*, without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. In no event shall the authors or copyright holders be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with
10 the software or the use or other dealings in the software." Please read carefully and abide by the full license before downloading and using this software.

Acknowledgements. The support of Prof. Barend Erasmus, Director of the University of the Witwatersrand's Global Change Institute (GCI) for the MISR-HR project since 2016 is greatly appreciated. Hugo De Lemos is financially supported through a grant by Exxaro Limited to the Exxaro Chair in Global Change and Sustainability Research. The computer used at GCI for processing MISR data was originally acquired
15 through grant No. 97853 from the South African National Research Foundation (NRF). Last but not least, the first author thanks Dr. David J. Diner of the NASA Jet Propulsion Laboratory (JPL), Principal Investigator of the MISR instrument, for his continuing support over the last 20+ years.

References

Bruegge, C. J., Diner, D. J., Korechoff, R. P., and Lee, M.: Level 1 Radiance Scaling and Conditioning Algorithm Theoretical Basis, Tech. Rep. JPL D-11507, Revision E, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, https://eospso.gsfc.nasa.gov/atbd-category/45, 1999.

5 Bull, M., Matthews, J., McDonald, D., Menzies, A., Moroney, C., Mueller, K., Paradise, S., and Smyth, M.: Data Products Specifications, Tech. Rep. JPL D-13963, Revision S, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, https://eosweb.larc.nasa.gov/project/misr/dps, 2011.

Diner, D. J., Bruegge, C. J., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Sellers, P. J., Clark, J., Daniels, J. A., Danielson, E. D., Duval, V. G., Klaasen, K. P., Lilienthal, G. W., Nakamoto, D. I., Pagano, R. J., and Reilly, T. H.: MISR: A

10   Multiangle Imaging SpectroRadiometer for Geophysical and Climatological Research from EOS, IEEE Transactions on Geoscience and Remote Sensing, 27, 200–214, https://doi.org/10.1109/36.20299, 1989.

Diner, D. J., Beckert, J. C., Reilly, T. H., Ackerman, T. P., Bruegge, C. J., Conel, J. E., Davies, R., Gerstl, S. A. W., Gordon, H. R., Kahn, R. A., Martonchik, J. V., Muller, J.-P., Myneni, R. B., Pinty, B., Sellers, P. J., and Verstraete, M. M.: Multi-angle Imaging SpectroRadiometer (MISR) Instrument Description and Experiment Overview, IEEE Transactions on Geoscience and Remote Sensing, 36, 1072–1087,

15   https://doi.org/10.1109/36.700992, 1998.

Diner, D. J., Asner, G. P., Davies, R., Knyazikhin, Y., Muller, J.-P., Nolin, A. W., Pinty, B., Schaaf, C. B., and Stroeve, J.: New Directions in Earth Observing: Scientific Applications of Multiangle Remote Sensing, Bulletin of the American Meteorological Society, 80, 2209–2228, https://doi.org/10.1175/1520-0477(1999)080<2209:NDIEOS>2.0.CO;2, 1999a.

Diner, D. J., Di Girolamo, L., and Clothiaux, E. E.: Level 1 Cloud Detection Algorithm Theoretical Basis, Tech.

20   Rep. JPL D-13397, Revision B, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, https://doi.org/10.5067/Terra/MISR/MIRCCM_L2.004, https://eospso.gsfc.nasa.gov/atbd-category/45, 1999b.

Diner, D. J., Braswell, B. H., Davies, R., Gobron, N., Hu, J., Jin, Y., Kahn, R. A., Knyazikhin, Y., Loeb, N., Muller, J.-P., Nolin, A. W., Pinty, B., Schaaf, C. B., Seiz, G., and Stroeve, J.: The value of multiangle measurements for retrieving structurally and radiatively consistent properties of clouds, aerosols, and surfaces, Remote Sensing of Environment, 97, 495–518, https://doi.org/10.1016/j.rse.2005.06.006,

25   2005.

Jovanovic, V. M., Lewicki, S. A., Smyth, M. M., Zong, J., and Korechoff, R. P.: Level 1 Georectification and Registration Algorithm Theoretical Basis, Tech. Rep. JPL D-11532, Revision D, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, https://doi.org/10.5067/Terra/MISR/MI1B2T_L1.003, https://eospso.gsfc.nasa.gov/atbd-category/45, 1999.

Stubenrauch, C., Rossow, W., and Kinne, S.: Assessment of Global Cloud Datasets from Satellites, WCRP Report 23/2012, World Climate

30   Research Programme, Geneva, Switzerland, 2012.

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

Zhao, G. and Di Girolamo, L.: A cloud fraction versus view angle technique for automatic in-scene evaluation of the MISR cloud mask,

35   Journal of Applied Meteorology, 43, 860–869, https://doi.org/10.1175/1520-0450(2004)043<0860:ACFVVA>2.0.CO;2, 2004.