DEVELOPMENT AND VERIFICATION OF THE NASA MULTI-ANGLE IMAGER FOR AEROSOLS OPERATIONAL CLOUD MASK

BY

JAVIER ALFREDO VILLEGAS BRAVO

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Atmospheric Sciences
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Adviser:

Professor Larry Di Girolamo

ABSTRACT

The National Aeronautics and Space Administration (NASA) Multi-Angle Imager for Aerosols (MAIA) instrument is set to launch in 2022 with the mission of quantifying the epidemiological relationships between aerosols and human health. The MAIA instrument's primary product is a level 2 aerosol particulate matter concentration measurement collected over cloud free pixels. The quality of this product heavily depends on the validity of the cloud mask. In this project, we present a cloud masking algorithm for MAIA constrained to its hardware. It consists of 7 observables that are tested against predetermined static thresholds. Both observables and thresholds are a function of scene type, which is a unique combination of sun-view geometry, day of year and surface type, including a novel surface classification scheme derived from the Multi-Angle Implementation of Atmospheric Correction Bi-Directional Reflectance Distribution Function (MAIAC BRDF) data set. The cloud mask algorithm works by checking if an observation exceeds or falls short of a threshold for any of the 7 observables, resulting in a cloudy or clear classification. The thresholds are derived to match the performance of the Terra Moderate Resolution Imaging Spectro-Radiometer (MODIS) high-confidence-cloud cloud mask to achieve cloud conservative behavior. The algorithm allows tuning of the conservativeness by introducing the quantities of Distance-to-Threshold, Activation Value and number of tests to activate. These user specified parameters determine how much confidence is needed for a cloudy or clear classification. The results are presented for the Los Angeles primary target area. The overall agreement between the MODIS cloud mask and the MAIA cloud mask (MCM) is 92.9%. Of the 7.1% disagreement, 60% of it was due to false positives by the MCM, considering MODIS as the truth. The MCM is in more than 90% in agreement with MODIS for deep non-sun-glint water and the first 11 of the 16 snow-free land surface types. It differs from the MODIS cloud mask the most

over bright desert, mountains and coastlines due to false cloudy flags. It agrees well with the MODIS cloud mask for cumulus, stratus and high cirrus, with greater disagreements over cloud edges, smoke plumes from wildfires, and very thin cirrus. The MCM agrees well with the MODIS cloud mask (>85%) for most solar zenith angles between 25 and 53 degrees, viewing zenith angles less than 60 degrees, and relative azimuth angles between 105 and 135 degrees. Several recommendations for improving the MCM are discussed, and its advantages over the MODIS cloud mask.

ACKNOWLEDGEMENTS

I'd like to thank Prof. Larry Di Girolamo, Dr. Yizhe Zhan and Dr. Guangyu Zhao for their invaluable support that has made this thesis possible. Yizhe was responsible for creating and coding the land surface ID, Guangyu has produced the thresholds and the surface ID files for operational use and helped me edit this thesis, and Larry has provided great guidance to allocate resources and give a focused vision of the MAIA cloud mask. I would also like to thank Larry and Guangyu for their encouragement and candid advice that has kept me sane during the writing process.

Additionally, I would like to thank my wife, Brooke Villegas. She has been there from the start because she truly believes in me and what I invest myself into. I am forever grateful to her for that.

I'd also like to thank my family. I have received nothing but support from them and am so thankful to have them by my side. They have pushed me so hard, because they wanted me to be the best person I could be.

Finally, I would like to thank the University of Illinois at Urbana-Champaign Department of Atmospheric Science. I have been at UIUC since 2015 as an undergraduate student and can say with certainty that this is the greatest department on campus. I have enjoyed the learning and the community so much. I am excited to be staying in the field to see everyone at conferences!

TABLE OF CONTENTS

CHAPTER 1 : INTRODUCTION

1.1. MAIA MISSION BACKGROUND

The Global Burden of Disease found that in 2016, 4.09 million deaths could be attributed to air pollution, and that air pollution ranked 7th for disability adjusted life-years (DALYs), which is the sum of years of life lost due to severe disability and premature death (GBD, 2017). One reason air pollution is so detrimental to human health can be explained by its particulate matter (PM) component, which varies in size and chemical composition. PM consists of particles suspended in the air made from a mixture of liquids and/or solids (WHO, 2013). There are two size ranges of relevance to human health. PM 10 refers to particles 10 μm or less in diameter which can be inhaled by humans and PM 2.5 refers to particles 2.5 μm or less in diameter which are able to reach much deeper into the human respiratory system (WHO, 2013). Depending on composition and length of exposure to the PM, it can cause many health problems including acute respiratory infections, chronic obstructive pulmonary diseases, and it is strongly associated with cancer and cardiovascular issues such as strokes and ischaemic heart disease (WHO, 2013).

PM air pollution originates from sources such as car exhaust, factories, biomass burning, wildfires and desert dust (Chow, 1995). It may be released and inhaled directly or react with other gasses in the atmosphere causing harmful substances to form (WHO, 2013). Chow (1995) and citations within show the chemical composition of the PM of interest includes elemental carbon, organic carbon, sulfate, nitrate and ammonium and that tracking air pollution and knowing its composition is paramount to performing attribution to human health and later dealing with those pollution sources. Brimblecombe (1987) showed one of the earliest recorded air pollution regulations from 14th century England, which was due to a respiratory illness that was traced from

black dust deposits on buildings and reduced visibility to an increased burning of coal. The capability of measuring these relationships beyond crude observation has the ability to improve quality of life on much large scales. Satellite remote sensing missions like the Multi-Angle Imager for Aerosols (MAIA) are seminal to improving the understanding of the relationship between air pollution and human health.

Diner et al. (2018) overviews the MAIA mission in great detail. They show the need for a passive remote sensing satellite to fill in the data gaps that expensive and sparse ground monitors leave, and the improvements the mission will have over MISR, the Multi-Angle Imaging Spectro-Radiometer on board the Terra satellite (Diner et al., 1998). The Terra satellite will be decommissioned to a lower orbit over the next couple of years due to low fuel and the mission will come to end (Kelly et al., 2014). MAIA as the successor will offer many improvements over MISR, including 10 more channels, 3 of which are polarized, and 1km resolution compared to MISR's 1.1km resolution. This allows MAIA to observe more particle properties like speciation, particle size and spatial distribution (Diner et al., 2018). **Table 1** shows MAIA's spectral specifications. MAIA's operational products are shown in **Table 2**.

Table 1: MAIA Band Specifications *This O2-A band is for research study use and excluded in operational aerosol retrieval. Table 2 from MAIA cloud mask algorithm theoretical thresholds document JPL-103722, Di Girolamo et al. (2019).

Band number	Band center (nm)	Bandwidth (nm)	Polarization	Main purposes
1	365	37		Aerosol absorption and height
2	391	39		Aerosol absorption and height
3	415	39		Aerosol absorption and height
4	444	53	Yes	Aerosol absorption and height, fine mode aerosol size, aerosol refractive index; cloud masking
5	550	43		Fine mode aerosol size; surface reflection; cloud masking
6	646	72	Yes	Fine mode aerosol size; aerosol refractive index; cloud masking
7	750	18		Fine mode aerosol size, bracket absorption bands
8	763*	6		Aerosol and cloud height detection
9	866	52		Fine mode aerosol size, bracket absorption band; surface reflection; cloud masking
10	943	46		Water vapor abundance
11	1044	97	Yes	Aerosol refractive index, bracket absorption bands, coarse mode aerosol size
12	1610	73		Coarse mode aerosol size; cloud masking
13	1886	83		Water vapor abundance; cloud masking
14	2126	114		Coarse mode aerosol size; surface reflection

Table 2: MAIA operational products. Figure from MAIA JPL website: https://maia.jpl.nasa.gov/resources/data-and-applications/

Name	Description
Level 1 Radiance and polarization imagery	Calibrated and georectified Stokes parameters describing radiance and linear polarization; view and solar geometry; latitude and longitude
Level 2 Cloud-screened aerosol properties	Cloud-screened total and fractional aerosol particle properties at time of satellite overpass
Level 2 24-hour averaged PM concentrations on overpass days	24-hr averaged concentrations of coarse PM, fine PM, and fine PM components on days and locations coincident with cloud-free and quality-controlled instrument observations of the MAIA PTAs
Level 4 Gap-filled, daily, 24-hour averaged PM concentrations	Spatially and temporally gap-filled 24-hour averaged concentrations of daily coarse PM, fine PM, and fine PM components over the MAIA PTAs

The MAIA mission consists of a one instrument satellite to measure PM concentrations and speciation over predetermined primary target areas (PTAs) shown in **Figure 1**. The PTAs (blue dots) are chosen to be in highly populated regions all over the world that have low frequency of clouds in order to maximize the amount of cloud-free aerosol data. At each PTA there may be grounded aerosol speciation monitors and/or PM mass concentration monitors. Additionally, there are secondary target areas (green dots) which are of scientific interest but not in the operational stream of the mission due to budget. The calibration target areas (orange dots) are chosen for "instrument calibration, stability monitoring and aerosol/PM validation" (Diner et al., 2018). All these components participate in 2 primary processes. MAIA has equipped one camera, in a step-and-stare mode, to scan each PTA at multiple angles. Simultaneously, measurements are being

taken by ground monitors. This provides a ground truth to back out the PM concentration and speciation from MAIA's Aerosol Optical Depth (AOD) over the entire swath. This is done with the use of geostatistical regression models. The second process involves epidemiologists analyzing health records on the ground along with MAIA's output to find a relationship between air pollution and human health. To execute this a cloud screening or cloud masking step is required, shown in **Table 2** along with the rest of the MAIA operational products.

Figure 1: MAIA Target Area Locations by City. Figure is taken from the MAIA JPL website: https://maia.jpl.nasa.gov/investigation/#target_areas

1.2. WHAT IS A CLOUD MASK AND WHAT IS IT USED FOR?

In satellite remote sensing a cloud mask classifies an image, pixel by pixel, as cloudy or not cloudy. It is an important step for most geophysical products since knowledge of what is being observed is required to back out its properties. This is because assumptions are made in the retrieval algorithms about the pixel's contents that are used to calculate quantities dependent on the presence or absence of clouds. For example, some surface products that require an unobstructed view of the surface or products that are not interested in cloud properties would be best retrieved over non-cloudy pixels, such as sea surface temperature (SST) derived from infrared measurements. If taken over cloud contaminated pixels, the SST would be severely underestimated (Koner et al., 2016). Likewise, cloud properties such as cloud optical depth (COD) and cloud droplet effective radius would be low biased if taken over clear contaminated cloudy pixels (Werner et al., 2018). It is therefore important to recognize that cloud masks should be application driven, to best serve the mission of the instrument as discussed in Yang and Di Girolamo (2008). The MAIA mission is specifically interested in AOD to back out near ground PM concentration and speciation, so its hardware is tuned for aerosol retrievals, with the knowledge in mind that to produce quality AOD retrievals a quality cloud mask is needed. Because the hardware and data processing are so expensive to the mission, however, tradeoffs are made. For example, MAIA does not have any thermal bands which are essential for cloud masking (Ackerman et al., 1998). However, MAIA is better equipped than MISR for cloud masking with a water vapor absorption band for cirrus detection (band 13) and a snow absorption band (band 12) for cloud detection over snow (Diner et al., 2018).

1.3. MAIA CLOUD MASKING OBJECTIVES

To meet the objectives of the MAIA mission a cloud mask is required to pass cloud free pixels to the aerosol products. This is because MAIA uses AOD, among other variables, to retrieve PM concentration and PM speciation (Diner et al., 2018). If the pixel is cloud contaminated, MAIA will overestimate AOD and therefore bias aerosol properties since clouds have COD ranging from ~0.4 to well over 100 (Chen et al., 2000), while aerosols have AOD on the order of 1 when very thick and hazardous to human health (Luo et al., 2014). The minimum delivery requirements to the NASA Jet Propulsion Laboratory (JPL) are 1) a day-time cloud mask for each PTA, 2) the cloud mask is generated independently for each step-and-stare view of MAIA, 3) cloud conservative behavior and 4) tunable thresholds to adjust its cloud/clear conservative behavior.

The MCM's default behavior is to be cloud conservative, which refers to the mask labeling a pixel cloudy only if it is statistically likely to be a cloud (Yang and Di Girolamo, 2008; Jones et al., 2012). This is at the detriment of potentially mistaking a low confidence cloudy pixel, like a cloud edge or a thin cirrus cloud, as clear. This would result in a cloud contaminated sample being sent to the aerosol product, which will bias the PM retrievals. However, aerosol filled data points could be lost for the epidemiological team, looking for a link between health and air pollution if all low confidence cloudy pixels are called cloudy, since heavy pollution can sometimes appear like clouds. Because it is not feasible to make a perfect cloud mask, low confidence cloudy pixels are preferably classified as clear based on epidemiologist's expert opinion. By choosing this cloud conservative framework over a clear conservative one, the number of "cloud free" samples sent to the aerosol product is maximized. To address the potential cloud contamination, a confidence metric called "Distance to Threshold" (described in **Chapter 2.4**) was created to allow tuning of the cloud mask conservativeness, even after the thresholds are finalized. This way a scientist at a

later date can achieve a desired result based on the purpose of their study and the performance of the cloud mask.

1.4. CLOUD MASKING BACKGROUND

1.4.1. INTRODUCTION

The field of cloud detection started to mature in the 1980s with the simultaneous advancement of computers and an increase in satellite missions, which allowed the automatic classification of pixels to be possible (see review of cloud masking in Goodman and Henderson-Sellers, 1988). Cloud detection is done by examining observables, such as visible reflectance, to exploit the contrast between cloudy (bright) and non-cloudy (dark) signals. Choosing which observables to use is limited by the measurement capabilities of the satellite instrument. Once chosen, however, they are gathered into logical operations that classify the data. The logic can be black box, as is the case for deep learning algorithms, or it may employ the use of thresholds in a decision tree. The thresholding method is the most common. It divides data from an observable into cloudy and non-cloudy classes based on expert analysis or by automatic calculation. There are many ways to derive the thresholds, and while they are tuned to maximize performance, they must also take into account downstream products that will use its output. Presented next are several cloud detection algorithms to show the diversity of how this pattern recognition problem has been framed in the past.

ISCCP, the International Satellite Cloud Climatology Project (Schiffer and Rossow, 1983), aimed for the first time to create a unified dataset of satellite radiances across numerous missions in order to infer the global distribution of cloud radiative properties. The project's ultimate goal was to lead to an improved representation of clouds in climate models using that observational

data. To analyze the clouds, however, it first required a unified cloud mask that was very sensitive in order to detect low clouds and thin cirrus clouds thought to cause important climate feedbacks in global climate models (Rossow and Garder, 1993). The algorithm takes advantage of the differences in the space-time variability of visible and infra-red (IR) radiances in cloudy and non-cloudy scenes. It uses the fact that cloudy radiances will be colder and brighter than the clear sky radiances to detect clouds. The scheme initially uses hand chosen thresholds to distinguish between the cloudy and non-cloudy data as a function of the surface type, defined as ocean, ice-covered ocean and land to create a reference high-confidence clear sky dataset based only off the IR channel. The radiance data is then grouped by intervals of (long- and short-term) time and by a more in-depth surface classification scheme to account for weaknesses in the first step caused by fluctuations of space-time variability on scales longer than days and across all regions of the globe. Then long- and short-term statistics (i.e. brightness temperature (BT) maximums, BT averages, visible radiance clear sky modes) from the radiances are extracted to modify the initial hand chosen thresholds for the IR and visible tests. The modifications are to allow more or less data to be classified as cloudy after learning about a region's clear sky and cloudy characteristics. The exact modification of the thresholds is described in section 3, "Description of algorithm", subsections c and d of Rossow and Garder (1993).

Statistical techniques around that time were being developed to detect clouds as well, similar to Gaussian mixture models, detailed by Simmer et al. (1982) and Phulpin et al. (1983). In these methods, multi-dimensional histograms are built using radiance data from different bands, such as the visible and IR, to locate clusters corresponding to distinct elements like clouds and the surface. These clusters are defined by a fitted Gaussian curve. When new data falls into a particular cluster, it takes on that clusters class. Additionally, because the continuum of data is known over the entire

space, it is possible to calculate the probability that a pixel is cloudy. This type of algorithm is unsupervised, meaning it can be trained without prior knowledge of the pixel's cloudiness, however, the Gaussian model will not necessarily fit the data perfectly and is a limitation of the method. Although, Phulpin et al. (1983) found that 87% of its cases classified by their model had comparable quality to manual analysis.

Saunders and Kriebel (1988), with the Advanced Very High Resolution Radiometer (AVHRR), a predecessor to MODIS, formed the foundation for modern cloud detection with the histogram-threshold method detailed in **Chapter 1.4.3** and **1.4.4**. The Terra MISR cloud mask (Diner et al., 1999) and the Terra MODIS cloud mask (Ackerman et al., 1998 and 2010) build on this and ultimately lead to the main logic (**Chapter 2**) and threshold derivation (**Chapter 4**) used in the MCM. This class of cloud mask derives thresholds to fit the result of a pre-labeled dataset. This is therefore a fully supervised technique unlike the Gaussian models or the ISCCP cloud mask discussed.

More recently in the past 5 or 10 years, large improvements in computing capabilities, data availability and open-source software has allowed a renewed interest in machine learning detailed by Jordan and Mitchell (2015). Machine learning offers the capability of forming complex non-linear relationships between variables and the desired output with automatically derived parameters. Furthermore, the rise of Python and open-source packages available for the public makes machine learning very accessible. There are community moderated libraries with every machine learning model imaginable written and debugged. Some notable libraries are Tensorflow developed by Google, pyTorch developed by the Facebook Artificial Intelligence (AI) Research Lab (FAIR), Sci-Kit Learn developed by the National Institute for Research in Computer Science and Automation (INRIA) and Caffe developed by Berkeley AI Research

(BAIR). Many of the models can run in minutes to hours on laptops (for simple datasets) and they can be trained with any data. The validation and testing procedures are standardized. What's missing is the domain expertise in the cloud masking community. Recently (White et al., 2020 in preprint) developed a Visible-Infrared Imaging Radiometer Suite (VIIRS) neural network cloud mask outperforming the continuity MODIS-VIIRS cloud mask (MVCM) (Frey et al., 2020) and National Oceanic and Atmospheric Administration's (NOAA) Enterprise Cloud Mask (ECM) as compared to collocated Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) observations. Wang et al. (2020) shows comparable performance for cloud conservative cloud detection on Suomi NPP VIIRS collocated with CALIOP using random forest models. Despite neural networks having been around for decades, however, and the recent innovations, machine learning has still not proven to be useful in operational cloud masking as compared to their traditional counter parts. Thus, for mission viability, the decision tree histogram approach for the MCM is chosen based on the MISR and MODIS cloud masks.

1.4.2. PASSIVE INSTRUMENT CLOUD MASKING

Although active instruments are the gold standard in cloud detection (i.e. Holz et al., 2008; Frey 2020 and citations within) because they can directly measure the backscatter of cloud particles and the retrieval height, they have very narrow swaths which reduces their ability to observe large regions of interest at one time, and subsequently have lower revisitation rates in orbit. For example, CALIOP, an active instrument, has a 70 meter wide footprint, versus Terra MODIS, a passive instrument, which has a 2030 km wide swath. Therefore, when wider swath missions are required, such as for MAIA's aerosol studies, a passive instrument is better suited to the task. Passive instruments measure short wave (SW) and long wave (LW) radiation reflected and emitted, respectively, by the atmosphere, the surface, aerosols and clouds. The unit commonly used is

radiance, which is the power per unit area incident on the sensor from a particular direction defined by the solid angle or instantaneous field of view (IFOV) of the camera. The SW radiances (<5 µm) mostly originate from the Sun, which varies in power output according to the wavelength, the Earth-Sun distance and the solar zenith angel (SZA) of the observation. SW radiance measured at the sensor, L_λ, at some wavelength, λ, is often converted to the at-sensor bidirectional reflectance factor (BRF) by normalizing the radiance by the cos(SZA) and the power output from the sun, E_λ, at a distance, *d*, from the sun as shown in **eq. (1)**.

$$BRF(\lambda\mu m) = \frac{\pi * L_\lambda * d^2}{cos(SZA) * E_\lambda} \qquad \textbf{eq. (1)}$$

The practical use of this specific definition of reflectance is that it is a directional point measurement from an infinitesimal solid angle, unlike albedo, which helps to capture the anisotropic characteristics of clouds and non-clouds, and it is normalized by the power output of the sun, unlike radiance. This definition follows from BRDF (bi-directional distribution function) as defined in Di Girolamo (2003), but is only different by a factor of pi when incident radiation comes from one direction at the altitude of the sensor. The LW radiances (>5µm) are emitted according to the wavelength, the temperature and the emittances of the elements in the pixel. For cloud masking these LW radiances are often converted to BT for thresholding, since clouds are typically colder than the surface. However, since MAIA does not have LW channels on board this is not further discussed. The SW BRFs are used to construct the observables for the MCM, detailed in **Chapter 2.3**, that are then thresholded to distinguish cloudy pixels from non-cloudy pixels.

1.4.3. CHOOSING THRESHOLDS FOR A CLOUD MASKING ALGORITHM

Cloud masks generally use observed or modeled data to train or select the parameters for the algorithm. Although it is possible to use a radiative transfer model to create synthetic data, it is often more straight forward and more representative to use real data to craft the thresholds. Despite knowing the truth perfectly in modeled data, it would be hard to capture the real variance in cloud types, cloud distributions, surface types, atmospheric profiles and so on to produce valid thresholds for each test. Although, efforts have been made to do just this, such as the global synthetic dataset or GSD (Grano et al., 2004; Shoucri and Hauss, 2009), which was deemed useful and within the cloud masking requirements for the VIIRS pre-launch thresholds for Suomi NPP (Hutchison et al., 2012). In both options, modeled or observed, a histogram can be built of the data for a particular observable to calculate the threshold at which a pixel is called cloudy. That data can either 1) be manually classified by an expert, 2) be previously classified by another algorithm, 3) be modeled synthetically so the absolute truth is known, or 4) a combination of the previous options. In any case, a high standard of ground truth is needed to serve as labels for the data. The quality of the cloud mask is function of the quality of the training data and its labels. First the data for one observable can be plotted on a histogram. If the tests are chosen correctly, a cloudy hump of data will appear distinct to the rest. Saunders and Kriebel (1988) first demonstrate the histogram approach in the most basic sense for one scene, shown in **Figure 2**.

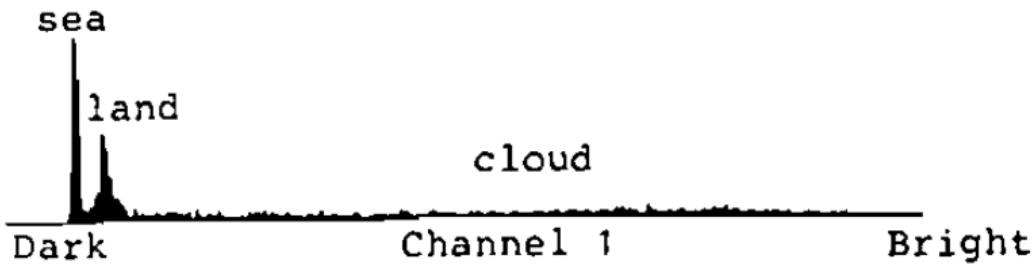

Figure 2: NOAA-9 AVHRR 50x50 pixel scene over the U.K. on April 15th, 1985. Figure from Saunders and Kriebel (1988).

Here a distinct signal in channel 1 visible reflectance for the NOAA-9 AVHRR is shown for a scene. On the left, or the dark reflectance, two peaks in the histogram correspond to sea and land. A threshold can be drawn to separate most of the observations with a level of certainty. To further distinguish the signal of cloudy pixels to non-cloudy pixels prior information about the measurement can also be incorporated. Meaning the histograms can be generated individually for unique combinations of sun-view geometries, surface types and seasons.

The spectral properties of different surfaces exist on a spectrum that approaches and overlaps with the spectral properties of clouds. In the context of a single threshold, it is paramount to maximize the contrast between the cloudy and non-cloudy observations for a particular test. This is true for any classification problem. This way a threshold can be drawn to separate the 2 classes with the least amount of model training errors. The easiest scenes will have a threshold that is clearly able to separate the cloudy observations from the clear ones and the most difficult scenes will have strict thresholds that may let cloud contaminated pixels be called clear. By dividing the histograms by scene type, the performance of the thresholds can be maximized in all cases. For example, cloud detection over deep ocean will have a much smaller region of ambiguity where the surface and cloud spectral properties overlap since water (non-sun-glint) is generally very dark and clouds are bright in the visible channel, versus cloud detection over a desert pixel where both cloud and desert are very reflective in the visible channels. It follows that the threshold over deep ocean will be much lower than the threshold over desert for a visible channel test. Therefore, the histograms are divided by surface type (surface type map defined in **Chapter 3**). The data can be can further stratified by the sun-view geometry. Minnis (1989) and Zhao and Di Girolamo (2004) for example, show that as viewing zenith angle (VZA) gets larger, cloud amount increases due to longer path lengths intercepting more cloud and larger GIFOV (ground instantaneous field of

views). Loeb et al. (1997) shows that retrieved COD increases for near nadir VZA as the SZA increases, since more solar radiation intercepts the cloud edges and then leaves the cloud top therefore appearing brighter. Liang and Di Girolamo (2013) show that, relative to nadir VZA, retrieved COD increases with increasing VZA in the backscatter direction and retrieved COD strongly decreases with VZA for the forward scattering direction. Therefore, cloud reflectance is a function of SZA, VZA and relative azimuth angle (RAA). Furthermore, the histograms can be divided by season. Surface types are dependent on the meteorology which can turn vegetation into bare soil, desert mountains into snow topped peaks or dry lake beds into deep water. As the surface reflectance changes, so must the thresholds in order prevent calling the newly bright surfaces cloudy or mistaking cloud contaminated data for clear as the surface gets darker. Moreover, cloud thickness and distributions change with the seasons and with latitude (Zhang et al., 2005). By changing the thresholds as a function of time and target area (region), a cloud mask can better distinguish cloudy and clear pixels. Given enough data and the right tests, the histogram can be divided into cloudy and non-cloudy sections with a threshold to a degree of certainty. Terra MISR and Terra MODIS end up taking a similar approach with these histograms and are the fundamental base for the MCM logic and threshold development.

1.4.4. TERRA MODIS MOD35 CLOUD MASK OVERVIEW

Ackerman et al. (2010) documents the theoretical basis for the latest MOD35 collection, 6.1. Most of the tests for MOD35 depend on mid IR and thermal emissive bands, but because MAIA's longest wavelength is 2126nm, these are not used in the MCM. Additionally, because the MCM is only operational during daytime, the MOD35 nighttime tests are not discussed. The like tests used in both MOD35 and the MCM reduce to the 0.65μm, 0.86μm and 1.38μm reflectance tests as well as a similar 3x3 windowed texture test, but instead of using the 11μm window channel,

MAIA will use the red band. Cloud reflectance is a weak function of wavelength in the visible and NIR part of the spectrum, giving clouds white color. Clouds are also generally brighter than the surface in the shortwave part of the spectrum. Over land the 0.65μm channel is used to detect clouds. Over water surfaces like the ocean, the 0.86μm channel is used instead, since water darkens into the NIR part of the spectrum, while clouds stay relatively bright (Ackerman et al., 1998). The 1.38μm channel is used for cirrus detection. Cirrus is usually optically thin and therefore washed out by the strong contribution of surface reflectance in channels where atmosphere absorption is small. At 1.38μm however, the surface is obscured by low level water vapor, which is highly absorbent at this wavelength (Gao, 1993). The contribution of radiance to the sensor is therefore predominately from above the moist layer. Here, high cirrus clouds stand out in contrast to the dark and opaque moist layer. MOD35 does not apply this test above 2000m surface elevation since mountains can peak above the moist layer and contribute large radiances to the sensor. Finally, at 1km resolution cloud tops and cloud edges are generally more textured than the surface except at surface boundaries like the coast and over mountains which are highly textured at 1km resolution. This test has heritage for MISR and MODIS over water, but it is also employed over land for the MCM since the MCM has a thorough thresholding calculation that takes into account surface types.

All these tests use simple thresholds dependent on surface types like land, water, and sun-glint water. The visible reflectance test however, used over land in collection 6.1 of the MOD35 cloud masking product, uses the histogram approach to produce thresholds dependent on the normalized vegetation index (NDVI) and the scattering angle. Each histogram is built using clear-sky/snow-free data from Aqua MODIS with one NDVI value, binned from 0 to 1 every 0.1 on 16-day intervals (Moody et al., 2005), and a scattering angle range. **Figure 3** shows the histogram for

Aqua MODIS visible reflectance band data from 2006 to 2007. The data all has NDVI between 0.7 and 0.8 and scattering angles between 110 and 120 degrees. The labels for the histograms come from the Aqua MODIS cloud mask record. In this MOD35 method, three thresholds are chosen as a function of NDVI (10 bins total from 0 to 1) and scattering angle. One to minimize error for a cloud conservative mask (red vertical line on right), one to minimize error for a clear conservative mask (blue vertical line on left), and the third is simply an average of the previous two (**Figure 3**). Therefore, four confidence levels are given: cloudy, uncertain clear, probably clear, and clear. A similar approach is used for all of MAIA's cloud masking observables (discussed in **Chapter 1.5**).

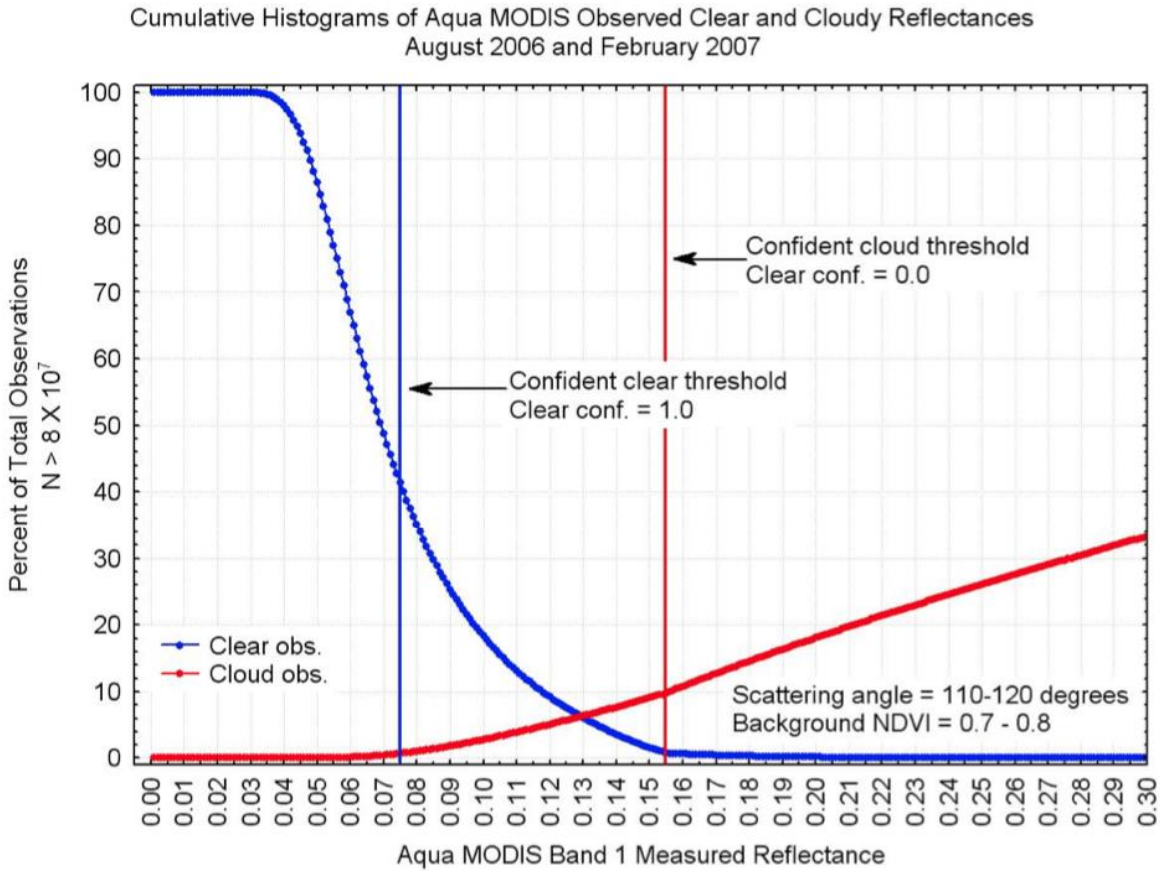

Figure 3: Clear (blue) and cloudy (red) cumulative histograms of 0.65µm Aqua MODIS reflectances from which a confident clear threshold (vertical blue line) and a confident cloudy threshold (vertical red line) may be defined. This is figure 8 from Ackerman et al. (2010).

1.4.5. TERRA MISR RCCM CLOUD MASK OVERVIEW

Terra MISR (Diner et al., 1998), the predecessor of MAIA (Diner et al., 2018), has a very similar cloud mask to the MCM described in Diner et al. (1999) called the radiometric camera-by-camera cloud mask (RCCM). Onboard, MISR has 4 channels: blue (446nm), green (558nm), red (672nm) and near IR (NIR) (867nm) with 9 cameras fixed at 9 different VZAs in the along-track direction. It borrows the histogram method and builds them for four observables and for different surface types, seasons and sun view geometries. This creates many thousands of thresholds that are automatically derived with the histogram technique, using semi manually classified scenes as the truth (Yang et al., 2007). The RCCM uses 4 observables, 2 over water and 2 over land. Over water RCCM uses the NIR BRF for high contrast with clouds and it uses a spatial variability index (SVI) since clouds are more textured than the ocean at 1.1km resolution. Over land the MISR RCCM builds histograms for each of the 1,580 surface types derived from the Cloud Screening Surface Classification (CSSC) dataset (NOAA-EPA Global Ecosystems Database Project) for 2 observables, D dependent on the NDVI and the red band BRF and SVI. Where NDVI is high for vegetation but lower for clouds (Zhu and Woodcock, 2012). The histograms for the land type are divided into 16-day intervals, 9 VZA bins for each camera, 10 SZA bins and 12 RAA bins. For each histogram, MISR has 3 thresholds to produce 4 confidence levels of cloudiness just like MOD35 (see **Figure 4**).

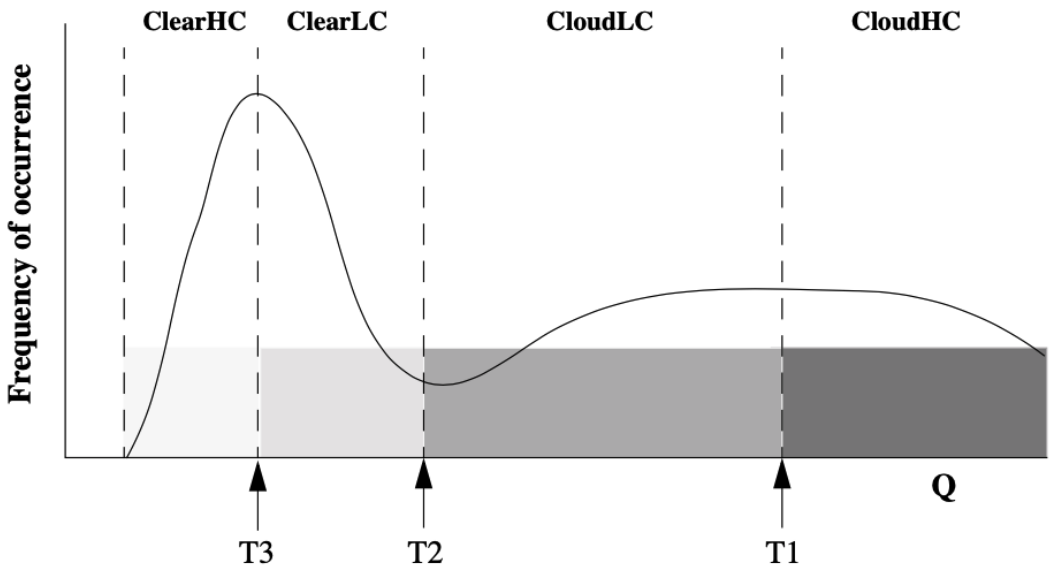

Figure 4: Cartoon of MISR histogram approach. HC is high confidence, LC is low confidence. Figure 1 from Diner et al. (1999)

1.4.6. CLASSICAL ISSUES IN CLOUD MASKING

There are several common issues faced in cloud masking. Surface reflectance can be just as bright as clouds over snow and sun-glint water, surface temperatures can be just as cold as clouds for snow-ice or within a temperature inversion, and surface textures can be just as high as clouds for surface boundaries and over mountains (i.e., Platnick et al., 2003; Ackerman et al., 1998). Furthermore, optically thin clouds and sub-pixel clouds are difficult to detect because of the contributions of radiance measured at the sensor come from both the cloud and the underlying surface (i.e., Weilicki and Parker 1992; Zhao Di Girolamo 2006). And finally, the presence of thick aerosol can contaminate the cloud mask due to the high reflectivity, similar to that of clouds (i.e., Lyapustin and Frey 2008). Thresholding these physical quantities as a function of the scene type is common to account for these exceptions. The surface types such as water, desert or snow present different backgrounds which change the detectability of clouds. The surface also tends to change with the seasons so thresholds can be dynamic over time (i.e., Ackerman et al., 1998; Diner

et al., 1999) and finally, the thresholds can be a function of sun-view geometry to take into account the angular scattering properties that distinguish clouds from the background (i.e., Minnis, 1989; Zhao and Di Girolamo, 2004).

1.5. CLOUD MASKING APPROACH FOR THE MCM

The MCM functions on 7 observables used to distinguish cloudy pixels from non-cloudy pixels. They are 0.65μm BRF, 0.86μm BRF, 1.38μm BRF, a texture index, a whiteness index, the normalized difference vegetation index and the normalized difference snow index, all discussed in detail in **Chapter 2.3**. The MCM logic depends on the use of thresholds for each test to classify the pixel. To calculate the thresholds, a labeled dataset is a required to build the histograms. In this thesis, I chose a proxy instrument to MAIA with a robust cloud mask to build the labeled dataset. The instrument must have a 1km resolution or better cloud mask, sufficiently similar radiance bands and sun-view geometry to MAIA, and a long record of observations. The Terra MODIS instrument, which launched in 1999, matches all these requirements. The MCM thresholds were derived for all tests similar to MISR/MODIS/AVHRR except it will use a novel surface ID (**Chapter 3**), it will use the complete sun-view geometry (cos(SZA), VZA, RAA), including more VZA bins than MISR, it chooses one cloud conservative threshold unlike both MISR and MODIS, and it is binned at a finer 8-day temporal resolution vs 16 days for MISR and MODIS. The exact threshold development is discussed in **Chapter 4** and in **Chapter 2** the forward pass of the algorithm is described.

CHAPTER 2 : MCM ALGORITHM

2.1. ALGORITHM OVERVIEW

The MCM algorithm takes radiances, sun-view geometry, a snow-ice mask, a land-water mask, and the threshold database as inputs. The output is a pixel-level binary cloud mask (See **Figure 6** for algorithm flowchart). It uses 7 observables that are grouped into 5 tests according to the physics being exploited. The 5 tests are grouped as 1) the 0.65µm BRF (Vis) and 0.86µm BRF (NIR) tests, 2) the 1.38µm BRF (Cirrus) test, 3) the spatial variability index (SVI) test, 4) the whiteness index (WI) test, and 5) the normalized difference index for snow (NDSI) and for vegetation (NDVI). The tests are applied according to the scene type. The scene type is defined by each unique combination of surface type, sun-view geometry and day of year (DOY) bins detailed in **Table 3**.

Table 3: Binning of data used to select a threshold. *Scene ID contains land surface ID, coast, water, sun-glint water, and snow-ice masks.

Observable Level Parameters	Range	# of bins	Bin Interval	Unit
cos(SZA)	[0,1]	10	0.1	[unitless]
VZA	[0,70]	14	5	[degrees]
RAA	[0,180]	12	15	[degrees]
Scene ID*	[0,19]	20	1	[unitless]
DOY	[1,365]	46	8	[days]

Because the BRF of different surfaces can differ by an order of magnitude (Jedlovec, 2009) (**Figure 5**) the observables are not applied to all surface types (**Table 4**). The tests are however

applied to all bins of sun-view geometry and all DOY bins. For each test there is a unique threshold for each scene type observed. In **Chapter 4** the threshold derivation is described.

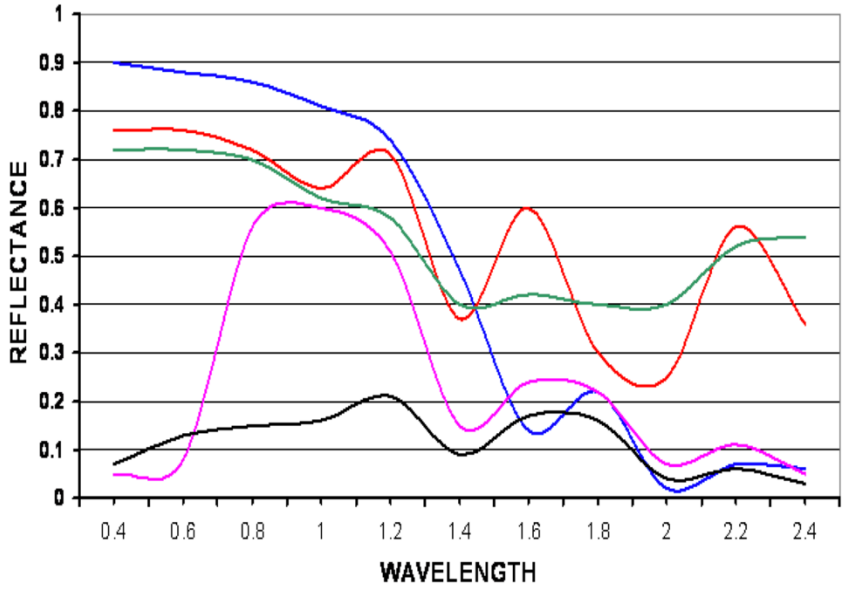

Figure 5: Typical reflectance values for snow (blue), bare soil (black), forest canopy (pink), cirrus (red), and stratus (green) clouds as a function of wavelength in μm. Figure from Jedlovec (2009).

Table 4: When each test is applied as a function of surface type. Note, all tests are applied for all sun-view geometry and DOY bins but is not shown here.

Test #	Observable	16 Land surface types + coast	water	sun-glint water	snow-ice
1	Vis Ref	yes			
1	NIR Ref		yes		
2	Cirrus	yes	yes	yes	yes
3	SVI	yes	yes	yes	yes
4	WI	yes	yes		
5	NDVI	yes	yes	yes	
5	NDSI				yes

The Next section describes the forward pass of the algorithm assuming the thresholds are already derived. This will show the basic logic and code of the MCM without yet introducing the most complex procedure in creating the MCM.

2.2. ALGORITHM FUNCTIONS

Figure 6 shows the input/output flow of the MCM code. On the left the inputs are color coated. The blue datasets are provided by the Di Girolamo group at the University of Illinois at Urbana-Champaign for operational use, while the red datasets are provided for by the NASA JPL (Jet Propulsion Laboratory). Because MAIA launches in 2022, this thesis uses the Terra MODIS instrument data as a proxy data for its similar channels, orbit and VZAs. The right side shows the functions (black elongated rhombuses) and datasets produced by the MCM (green elongated ovals). The surface ID derivation is discussed in **Chapter 3** and in **Chapter 4** the threshold derivation is discussed. The config file is described in the MAIA Cloud Mask ATBD (Di Girolamo et al., 2019). The functions shown in **Figure 6** are described in the rest of this section in the order from the top to the bottom of the flow chart.

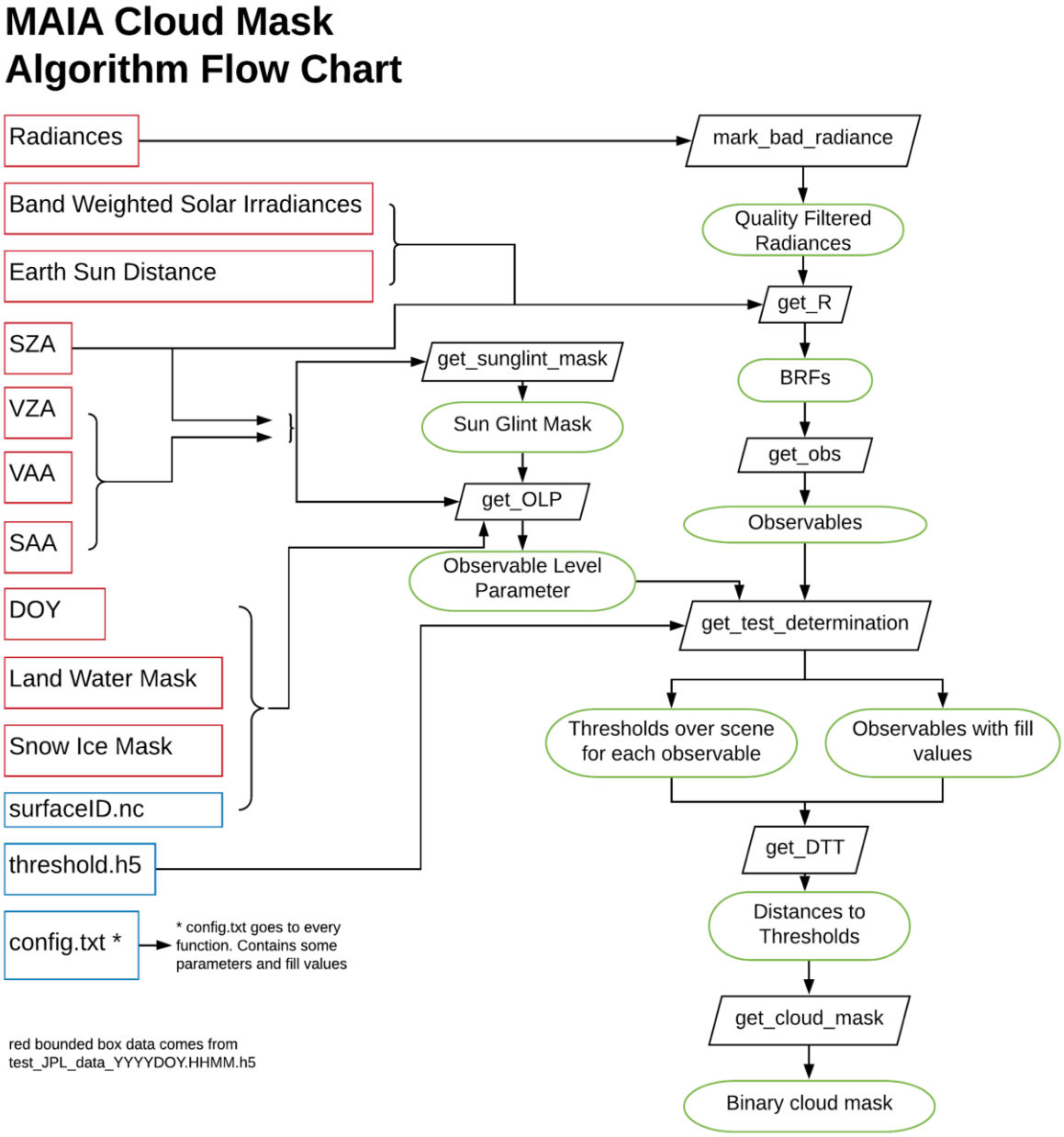

Figure 6: MCM algorithm flow chart. Red boxes are dynamic inputs provided by JPL at processing time. Blue boxes are static pre-launch inputs provided by UIUC. Black rhombuses are functions. Green ovals are input/output data into and from functions.

mark_bad_radiance()

The MCM will have a Radiance Data Quality Indicator (RDQI) mask from the Level 1 radiance product. The mask contains the integers from 0 to 3, where 0 is a high-quality retrieval, 1 is a low-

quality retrieval but still usable, 2 is a poor quality and unusable retrieval and 3 is no data. The MCM will only process the cloud mask when the required bands have 0 RDQI, although this can be changed in the configuration file. This function will mask the radiance data to -998 for RDQI > 0 and -999 for RDQI == 3. Data with RDQI == 0 will not be masked and contain the original radiance values.

get_R()

This function retrieves the Bi-Directional Reflectance Factor (BRF) for a given band. It uses the SZA, Earth-Sun distance and the band weighted solar irradiances. BRF for cos(SZA) <= 0.01 is assigned the fill value -998. Note that "R" is used interchangeably as a shorthand for BRF in equations. The equation for BRF is shown here again for convenience.

$$BRF(\lambda\mu m) = \frac{\pi * L_\lambda * d^2}{cos(SZA) * E_\lambda} \qquad \textbf{eq. (2)}$$

get_obs()

There are 7 observables which compose the 5 tests, which are all derived from 6 BRFs. For a scene, such as the 400km by 300km Los Angeles (LA) PTA, it would return the 7 observable values for each pixel. See **Table 4** for tests and see **Chapter 2.3** for more detail.

get_sunglint_mask()

$$\Theta = arcos(sin(VZA) * sin(SZA) * cos(VAA - SAA - 180°) + cos(VZA) * cos\,(SZA)) \qquad \textbf{eq. (3)}$$

eq. (3) calculates the scattering angle, Θ, using the SZA, VZA and RAA of the pixel, and then flags pixels with scattering angles between 0 and 40 degrees over water as sun-glint. 0 is sun-glint water, 1 is everything else. The range or cone of scattering angle, however, can be changed in the configuration file. The relative azimuth angle here is defined relative to Terra MODIS RAA = (VAA - SAA - 180) where the VAA and SAA are the viewing and solar azimuth angles. The forward scattering occurs at 0 degrees and back scattering occurs at 180 degrees. The sun-glint mask is integrated into the surface type mask.

get_OLP()

The observable level parameter (OLP) is the backbone of the MCM. It characterizes a pixel by its cos(SZA), VZA, RAA, Scene ID, DOY, and target area (TA). Each pixel has these prior knowledge data points that describe the scene in which a cloud may reside. The sun-view geometry is continuous from the data file while the Scene ID, DOY and TA are discrete values. The sun-view geometry is binned in this function to retrieve a valid threshold. **Table 3** shows how they are binned. The return of this function is an array the same length and width of the scene, and the depth has the binned values corresponding to each pixel. This will feed into get_test_determination() to pick out the thresholds for the scene.

get_test_determination()

Test determination serves two functions in 1. It masks the observables not used in the final MCM according to surface type (see **Table 4**). The second and primary function is to fetch the thresholds for the scene for each of the 7 observables. Each threshold is unique to the OLP of a pixel. An example of OLP for a pixel maybe [0,3,2,17,32,0]. This would represent cos(SZA) between 0 and 0.1, VZA between 15 and 20 degrees, RAA between 36 and 48 degrees, water surface type, DOY

between 257 and 264 inclusive in the Julian calendar and PTA 0 which is the LA PTA. These exact numbers from OLP are used to query the threshold database to retrieve one threshold for that pixel. This is done 7 times for each observable and for all the pixels in the scene. Therefore, this function returns a 2D array of thresholds for the subject scene, for each of the seven observables. The thresholds are then used to calculate the distance to thresholds.

get_DTT()

DTT or distance to threshold returns a confidence metric of cloudy or clear based on the thresholds for each observable. Where the test is not performed a fill value is applied. The calculations and reasonings are discussed later in **Chapter 2.4**. The DTT maps for each observable are fed to the next function to get the final cloud mask.

get_cloud_mask()

The final step is to use the DTT maps to determine which pixels were cloudy and which were not. DTT values of 0 or greater are considered to be cloudy. If at least one of the 7 observables meets this criterion, the pixel is determined to be cloudy. Although, the cloud mask configuration file has some options to control how many observables must show cloud and at what confidence it should be called cloudy. This is discussed in greater detail in **Chapter 2.4**.

On the Keeling cluster hosted at the University of Illinois' School of Earth Society and the Environment, the MCM forward pass runs in about 1.5 seconds for a 400 by 300 pixel scene on one Intel(R) Xeon(R) CPU model E5-2660 v3, clocked at 2.6 GHz with 4 Giga Bytes (GB) of memory. The input and output file sizes are all subject to change, so they are not listed. The next 2 sections describe the 5 tests and the logic used to determine if a pixel is cloudy or clear. Afterwards, the derivation of the thresholds is shown.

2.3. DEFINITION AND PHYSICS OF 5 TESTS

2.3.1. VISIBLE REFLECTANCE AND NIR REFLECTANCE TEST

In the visible and NIR wavelengths, clouds are generally brighter than the underlying surface as shown in **Figure 7**, where the Terra MODIS visible BRF and NIR BRF images for the LA PTA on Dec. 31, 2018 are displayed in gray scale (this scene is used for the rest of **Chapter 2.3**). The red band BRF is used to distinguish clouds from snow-ice-free land, since vegetation in the NIR are much brighter than in the Vis (**Figure 5**) and cloud BRFs are about the same in both channels. Since water is darker and atmospheric molecular scattering is lower in the NIR compared to Vis, 0.86μm BRF is used to distinguish clouds from sun-glint-free and ice-free water. The weaknesses for these tests are bright surfaces (other than water glint and snow-ice) like sandy coastlines, shallow water, arid soils and deserts. This is especially true for clear scenes and for optically thin clouds as they give strong contributions of reflectance from the surface to the sensor. If a cloud happens to escape detection by this test it is hoped to be found by the remaining tests as they exploit different cloud, atmosphere and surface properties.

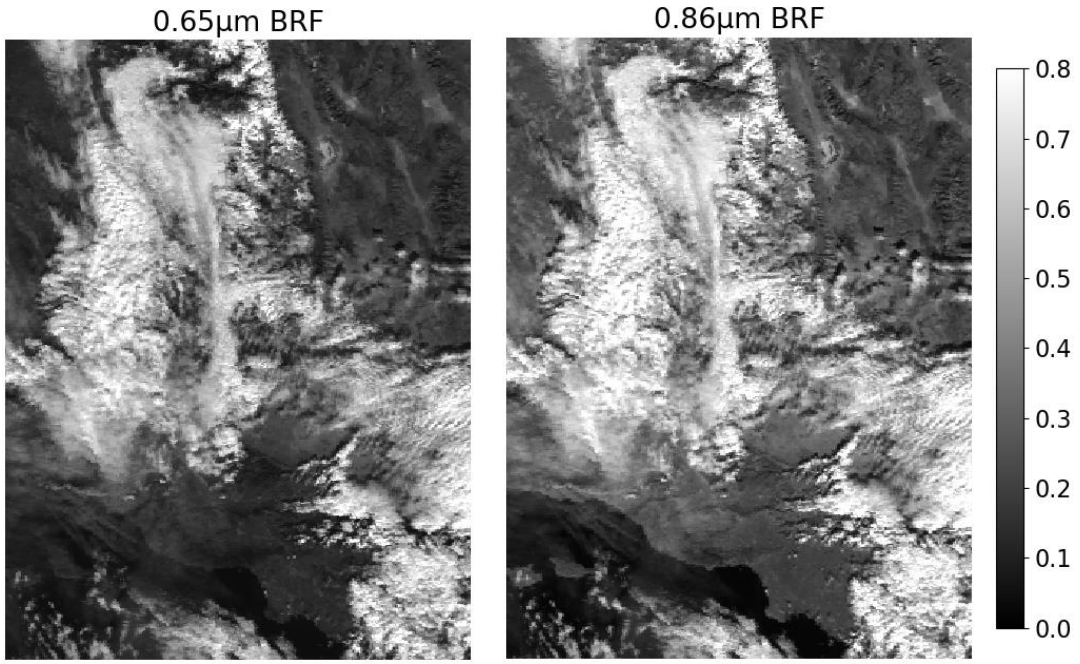

Figure 7: Visible BRF on left. NIR BRF on right. Both on same scale for Terra MODIS scene 2018365.1830 (<year><Julian day>.<UTC time>).

2.3.2. CIRRUS TEST

The cirrus test uses 1.38µm BRF to find clouds. This is a strong NIR water vapor absorption band. Since most water vapor in the atmosphere originates from the surface, and the boundary layer is usually no more than a few km deep, the surface is mostly opaque at this wavelength, and therefore very dark as seen in **Figure 8**, which shows visible (left) and cirrus (right) channels of Terra MODIS for the LA PTA on Dec. 31, 2018. The result is that the only contribution of reflectance comes from high clouds which generally exist in dryer environments giving stark contrast to the surface. Therefore, it has the potential ability to flag high thin cirrus that a visible band would not be as sensitive to as seen from comparing the visible (left) panel to the cirrus (right) panel in **Figure 8**. This test is very powerful, but it has 2 key weaknesses. It is not sensitive to low clouds, and it has the potential to flag high terrain above the moist layer as cloud. Ackerman et al. (2010) does not use the cirrus test on terrain over 2000m and should be implemented into the MCM at some

point. Moreover, the absence of clouds in the test does not mean clouds are not present. It relies on other tests to detect low clouds and clouds in very moist environments.

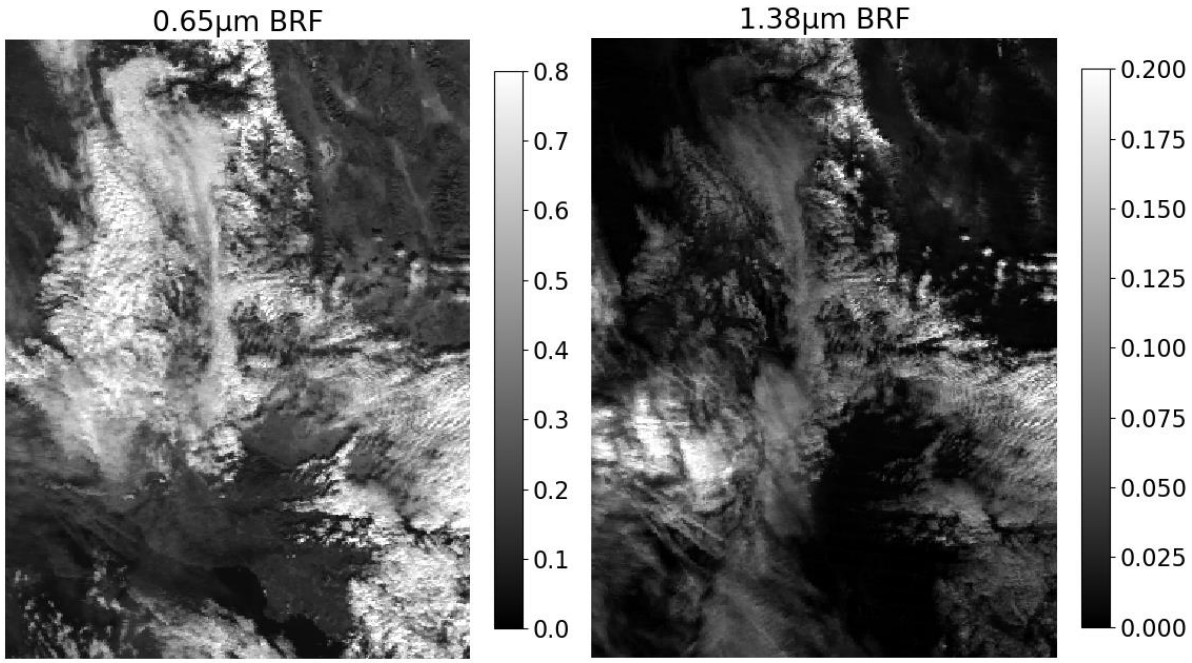

Figure 8: Visible BRF test on left. Cirrus test on right. Terra MODIS scene 2018365.1830 (<year><Julian day>.<UTC time>).

2.3.3. SPATIAL VARIABILITY INDEX (SVI)

At 1km resolution, clouds tops and cloud edges tend to have more texture than the surface. This is especially true over water. The SVI is applied over all surface types. The texture is calculated as the standard deviation of the red band BRF values over a sliding 3x3 window with a stride of 1 in all directions show in **eq. (4)**.

$$SVI = \frac{\sqrt{\sum_{i=1}^{9}\left(\bar{R}_{0.65\mu m} - R_{0.65\mu m_i}\right)^2}}{9} \qquad \textbf{eq. (4)}$$

The standard deviation is then mapped onto the pixel over which the window was centered as shown in Error! Reference source not found., with the visible channel on the left and the SVI on the right. If the pixel is at the edge of the image, the test is not applied. Any other boundary

condition would cause fake textures at the edge of the PTA that may or may not be consistent with the cloudiness of those pixels. Additionally, if a full 3x3 pixel window is not possible due to bad or missing data, the test is not applied since the definition of texture would change scale. That would require a separate threshold derivation and would suffer from low sampling issues.

This test has weaknesses over very textured terrain like mountains, which have large scale features similar to that of clouds. It also fails at surface boundaries, particularly at coastlines. Additionally, it is worth noting that some PTAs only have snow or ice in mountainous regions, which conflates the performance of this test over snow and ice with its performance over mountainous regions.

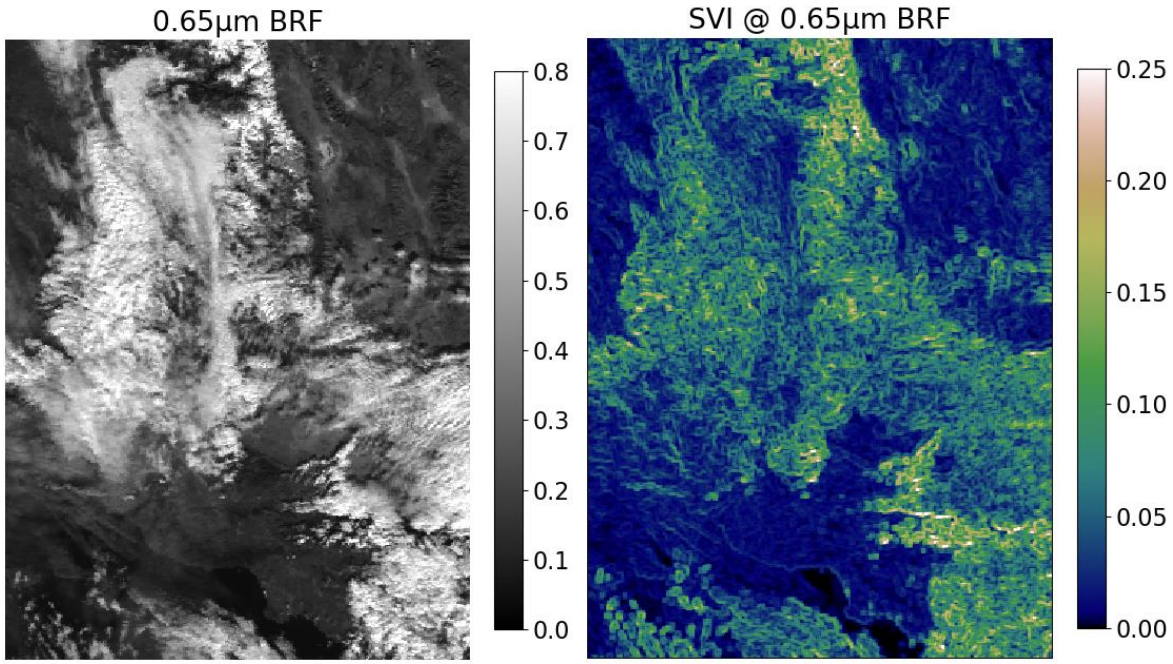

Figure 9: Visible BRF on left. SVI on right for Terra MODIS scene 2018365.1830 (<year><Julian day>.<UTC time>).

2.3.4. WHITENESS INDEX (WI)

The BRF of clouds is a weak function of wavelength in the visible and NIR parts of the spectrum (Zhu and Woodcock, 2012). This makes them white in color and therefore unique from most surfaces and aerosols. WI works by comparing each of the red, blue and green (RGB) bands' BRF (R_i) to the average RGB BRF ($\overline{R_{rgb}}$) for the pixel, summing the magnitude, and normalizing by the average BRF ($\overline{R_{rgb}}$) of the 3 colors as show in **eq. (5)**. The result is that a value of 0 corresponds to a perfectly white pixel, and more positive values are less white. The normalization makes the WI independent of brightness. Its main weakness is light colored surfaces like deserts and sandy coastlines when the scene is predominantly clear. It is not applied over sun-glint water regions at all.

$$WI = \frac{\sum_{i=1}^{3} |\overline{R_{rgb}} - R_i|}{\overline{R_{rgb}}} \qquad \textbf{eq. (5)}$$

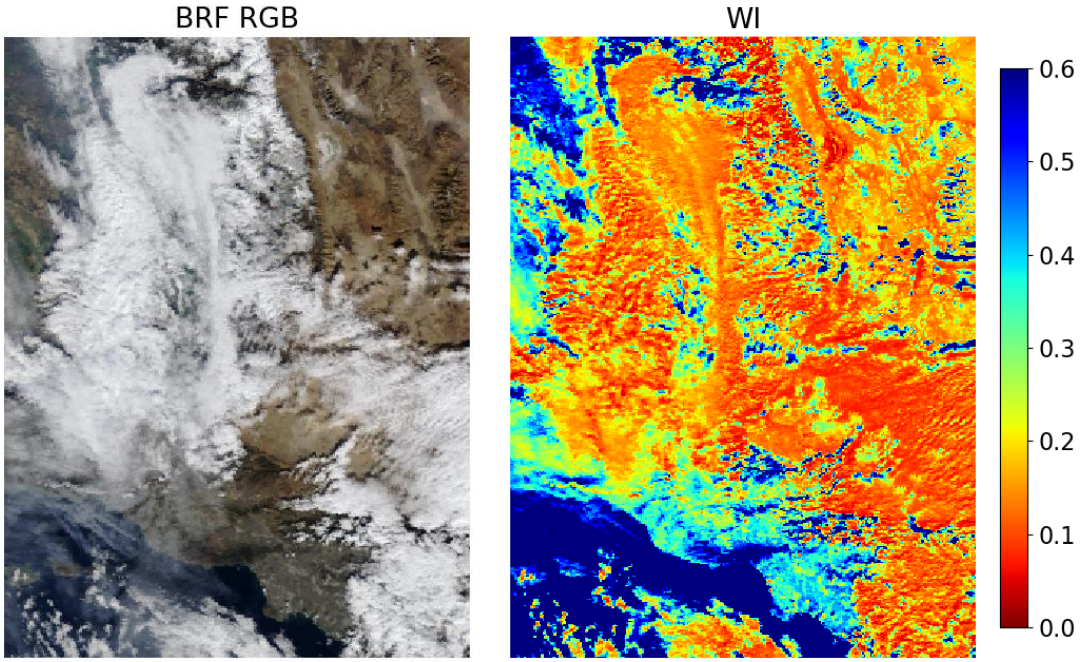

Figure 10: BRF RGB on left. WI on right. Terra MODIS scene 2018365.1830 (<year><Julian day>.<UTC time>).

2.3.5. NORMALIZED DIFFERENCE INDEX FOR SNOW AND VEGETATION

NDVI and NDSI take advantage of two things, the spectral properties of the subject, vegetation or snow, and the whiteness of clouds. They are defined below in **eq. (6)** and **eq. (7)**. In **Figure 11** an RGB of the LA PTA taken from Terra MODIS on Dec. 31, 2018 is shown alongside the NDVI and NDSI for that scene. Vegetation is much more reflective at 0.86μm than in the visible

$$NDVI = \frac{R_{0.86\mu m} - R_{0.65\mu m}}{R_{0.86\mu m} + R_{0.65\mu m}} \qquad \textbf{eq. (6)}$$

$$NDSI = \frac{R_{0.55\mu m} - R_{1.6\mu m}}{R_{0.55\mu m} + R_{1.6\mu m}} \qquad \textbf{eq. (7)}$$

so it has high values, while clouds tend to lower the NDVI towards zero (Zhu and Woodcock, 2012).

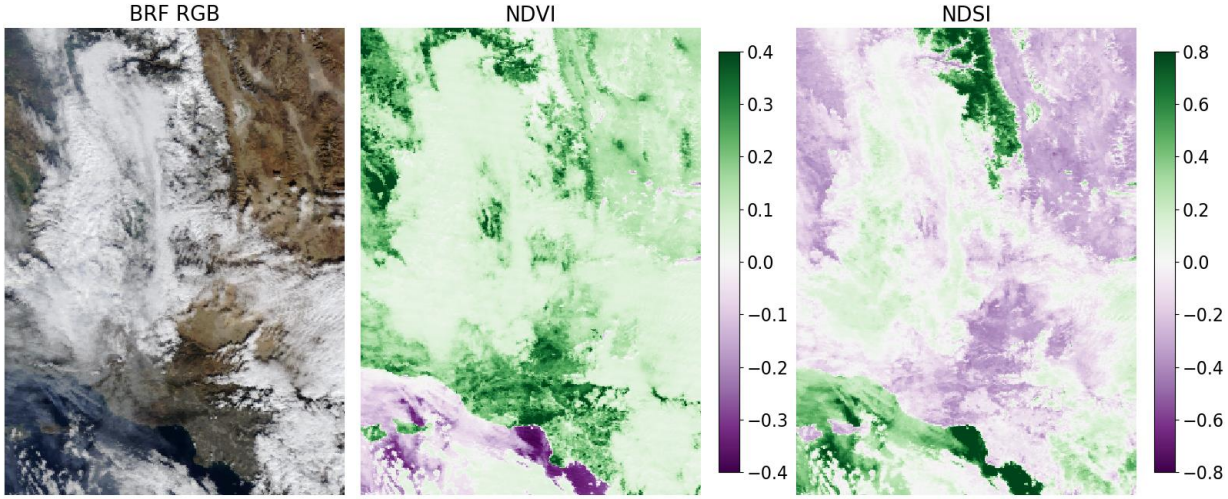

Figure 11: From left to Right: BRF RGB, NDVI, NDSI. Terra MODIS scene 2018365.1830 (<year><Julian day>.<UTC time>).

In the MCM, the NDVI is applied over all surface types except for snow-ice, where NDSI is applied instead. Snow is highly absorbent of 1.6μm radiation and highly reflective in the visible part of the spectrum, so snow takes on positive values generally > 0.4 (Hall et al. 1995, Dozier

1989), whereas clouds tend to reduce NDSI towards zero (Choi and Bindschadler, 2004; Warren 1982). Weaknesses of NDVI occur on coastlines and desert regions. These surface types share a very similar NDVI signature to clouds since they have very little vegetation and are close to white in color. NDSI and NDVI are also subject to at least the amount of error in the supporting snow-ice mask, since NDVI will call snow-ice a cloud, and NDSI will mistake water and arid land for cloud.

The next section uses the definition of the 5 tests to produce cloud confidence maps called distance to threshold that uses a simple logic to decide if each pixel is cloudy, clear or an invalid retrieval.

2.4. DISTANCE TO THRESHOLD (DTT), ACTIVATION VALUE (AV) AND THE FINAL CLOUD MASK

Distance to threshold or DTT is a metric used to quantify how much the threshold was exceeded by or fell short of an observable. This gives a proxy to confidence. There are 3 definitions of DTT, each one to accommodate the physical meaning of the different tests.

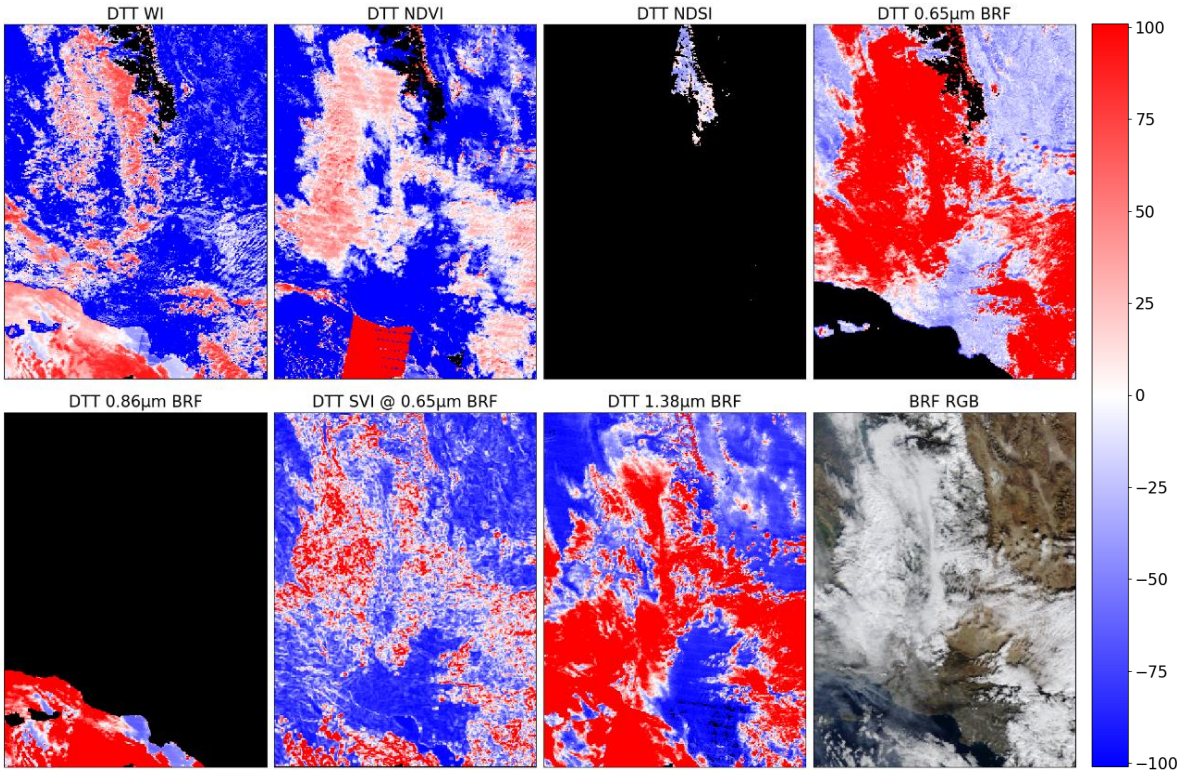

Figure 12: Distance to Threshold for all 7 observables for Terra MODIS 2018365.1830 (<Year><Julian DOY>.<UTC Time>).

$$DTT\ Ref = (R_\lambda - T)/T \qquad \textbf{eq. (8)}$$

$$DTT\ WI = (T - R_\lambda)/T \qquad \textbf{eq. (9)}$$

$$DTT\ NDxI = (T - |NDxI|)/T \qquad \textbf{eq. (10)}$$

Figure 12 shows the DTT for the 7 observables followed by an RGB image from the LA PTA taken from Terra MODIS on Dec. 31, 2018. The three definitions produce a map of the

original scene, where white blending into red (DTT>=0) denotes increasing cloudy confidence, while white blending into blue (DTT < 0) denotes increasing clear confidence. Black (DTT = fill values) denotes that the test is not applied over that surface type, bad data, or missing data to calculate the observable. Wherever any pixel has a value of DTT >= AV over at least N of the 5 tests, a final classification of cloudy is given. In the configuration file, the AV is 0 and N is 1. This means only 1 out of the 5 tests needs to exceed an AV of 0 to call the pixel cloudy, otherwise it is assumed to be clear if at least 1 test says so. The 0.65µm, 0.86µm and 1.38µm observables use **eq. (8)**, the WI uses **eq. (9)**, and NDSI and NDVI use **eq. (10)** to calculate the DTT. These definitions keep the definition of cloudy consistent for DTT >= AV. This is the last step before producing the final cloud mask shown in **Figure 13**.

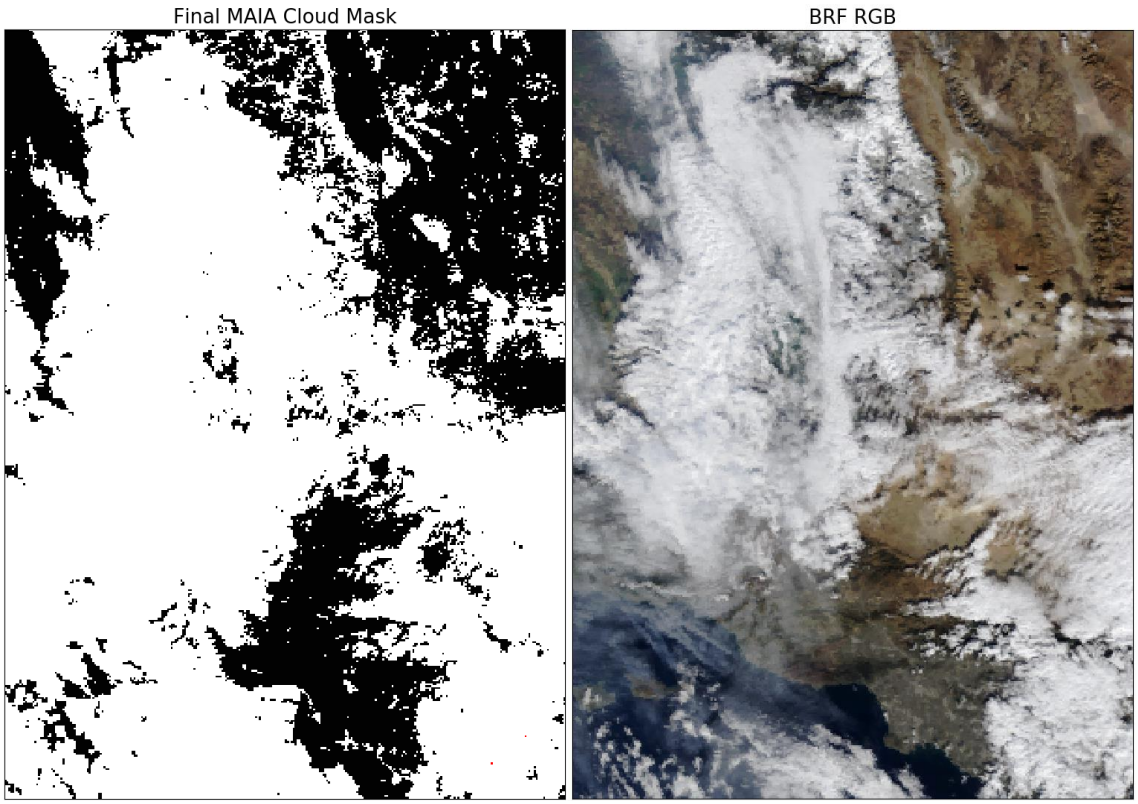

Figure 13: Final cloud mask and RGB BRF image. Terra MODIS 2018365.1830 (<Year><Julian DOY>.<UTC Time>).

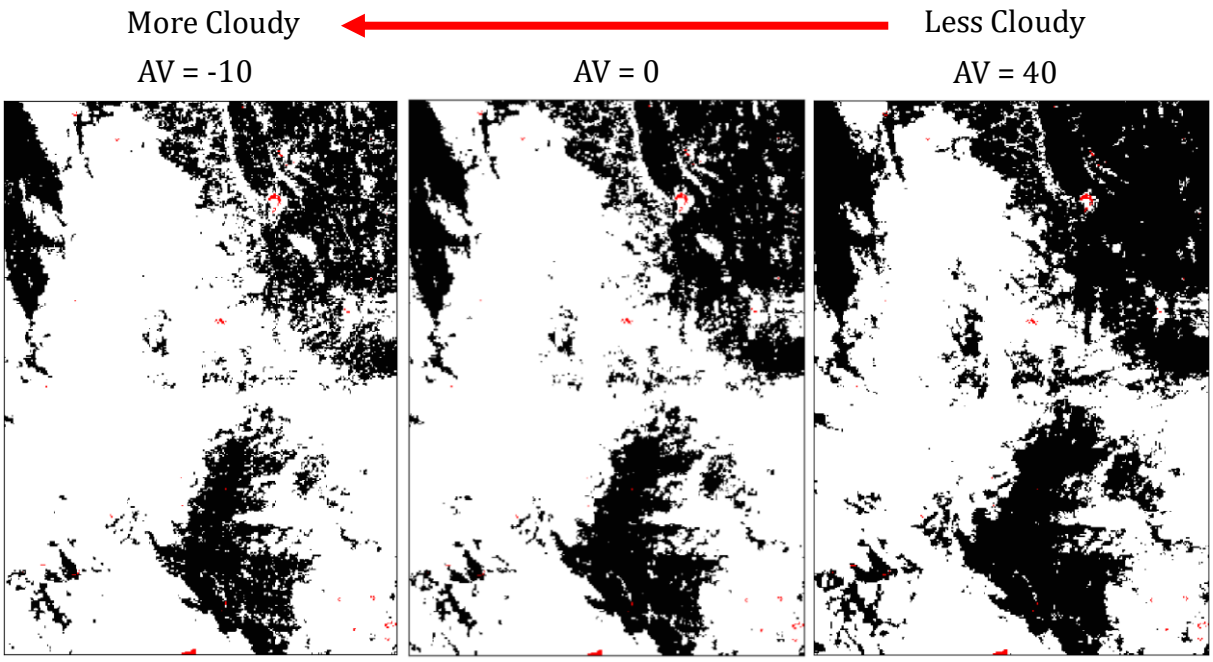

Figure 14: From right to left the AV decreases causing the cloudiness to increase denoted by the red arrow. Terra MODIS 2018365.1830 (<Year><Julian DOY>.<UTC Time>).

In **Figure 14** the cloud mask is shown again but with different AVs that can be changed in the configuration file. If the AV is changed from 0 to 40, the cloud fraction decreases, since the DTT must then be at least 40 for a pixel to be classified as cloudy. The DTT can also be lowered to negative 10 for example, and now the cloud fraction increases since a lower threshold for DTT is required for a pixel to be called cloudy. The biggest change occurs at the cloud edges and over the North East quadrant over the bright desert mountains. In other words, the areas of low confidence clouds can be masked out or reintroduced by the user according to the AV. This will allow flexibility for the downstream aerosol products.

The final output of the MCM is a designation of cloud, clear or unclassified pixels (**Table 5**). The projected output file datasets can be found in **Table 6**.

Table 5: MAIA Cloud Mask output

Output	**Integer Representation**
Cloudy	0
Clear	1
Bad Data	2
Missing Data	3

Table 6: MCM Projected Output File Contents

Data Set Name	**Description**	**Data Type**
Scene Type Identifier	sun-glint/snow-ice/land surface/water mask	32-bit integer
Final Cloud Mask	cloud mask	4-bit float
Distance to Threshold	confidence metric for each test	64-bit float
Observable Data	helps interpret DTT	64-bit float

This is the final product of the forward pass of the MCM. The algorithm rests on the threshold parameters to distinguish cloud and clear for the 7 observables. In **Chapter 3** the land surface ID generation is shown in support of the threshold generation shown in **Chapter 4** and in **Chapter 5** and **Chapter 6** the results and analysis are presented.

CHAPTER 3 : CUSTOM SURFACE ID

3.1. MOTIVATION

Due to the variability of the surface BRDF in space and time, a dynamic surface ID is crucial to having a good threshold dataset for a cloud mask. Many surface ID maps exist for cloud masking, but do not meet this higher standard. For example, MOD35 uses a 16-day interval map made from only a few years of NDVI data binned 10 time from 0 to 1 (Ackerman et al., 2010; Moody et al., 2005). The International Geosphere-Biosphere Programme (IGBP) is one NDVI based map fixed for the whole year from a couple of years of data (Loveland et al., 1997). The CSSC for MISR is also a static map (Diner et al., 1999). For the MCM we use the Multi-Angle Implementation of Atmospheric Correction (MAIAC) (Lyapustin et al., 2018) data to generate surface IDs. MAIAC provides a way to leverage 8-day BRDF datasets from Terra MODIS to classify the clear land surface by maximum observed reflectance. The land surface ID presented is static from year to year, but dynamic on an 8-day interval and at 1 km resolution. It is derived from 18 years of data, from 2002 to 2019, such that the extremes of the inter-annual variability in the recent past can be captured.

3.2. DATA: MAIAC

The MAIAC "MCD19A3" files contain three kernel coefficients for the Ross-Thick Li-Sparse (RTLS) equation to retrieve BRDF given the sun-view geometry (Lucht et al., 2000). The exact forms of the equations are given in Lucht et al. (2000) equation 37 for the final BRDF, equation 38 for the Ross-Thick kernel and equations 39-44 for the Li-Sparse kernel. The RTLS kernels depend only on sun-view geometry and are scaled by the three coefficients given in the MCD19A3 file. Retrievals of the three coefficients are only done over clear sky (no clouds or $AOD_{0.47\mu m} \geq$

1.5) and snow-free land. The coefficients are derived from only 3 observations of the same point belonging to one 8-day period as observed from Terra MODIS. These coefficients are provided for every pixel in a given MODIS tile (**Figure 15**) every 8-days starting Jan 8 at 1 km resolution and they vary from year to year for every year of the MODIS Terra mission.

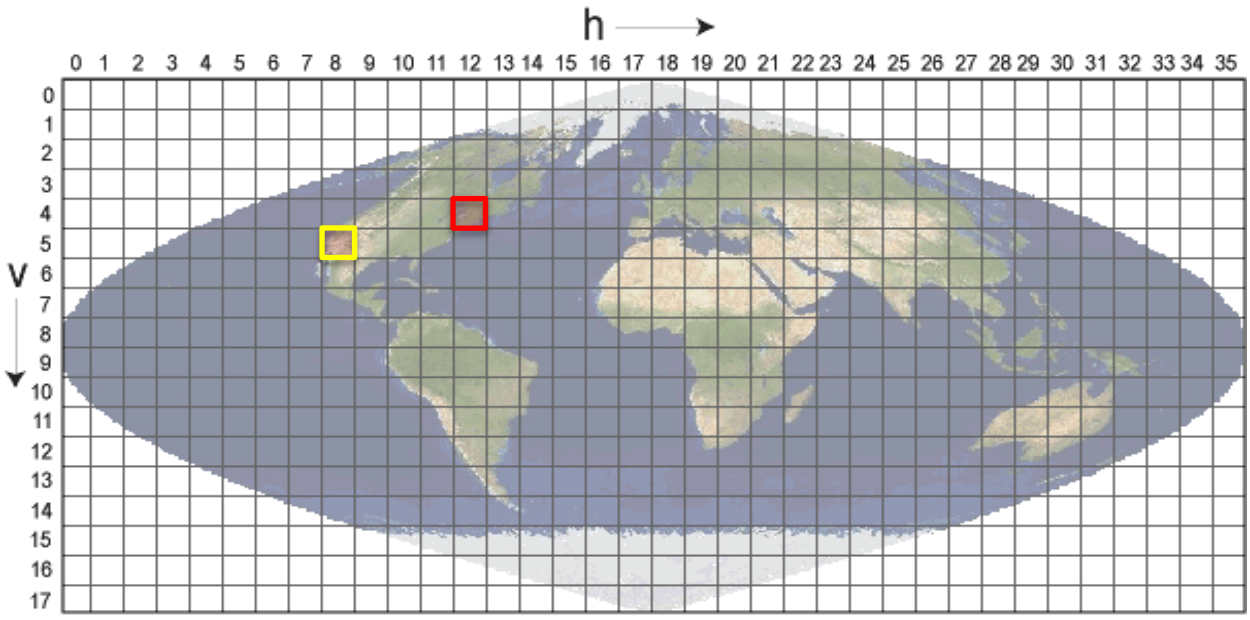

Figure 15: Map of MODIS sinusoidal tiles. Axis v and h are numbered and are used in MODIS products to select an area, i.e. "MCD19A3.A2019033.h12v04.006.2019042034207.hdf". Tile h12v04 is bordered in red and tile h08v05 is bordered in yellow. Figure taken from website: https://modis-land.gsfc.nasa.gov/MODLAND_grid.html

For the LA PTA the MCD19A3 RTLS kernels (**Figure 16**) from the h08v05 tile (bordered in yellow, **Figure 15**) is regridded to the PTA's bounding box defined by JPL shown in **Figure 17**, for years 2002 to 2019.

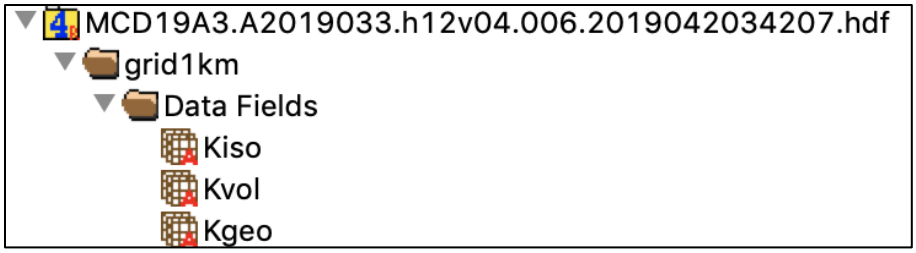

Figure 16: Example file containing 3 kernels for all pixels in tile h12v04 at 1km resolution viewed on HDFView.

Figure 17: LA PTA bounding box from JPL KMZ file viewed on Google Earth.

3.3. MAIAC DATA PRE-PROCESSESING AND ALGORITHM INPUT/OUTPUT

1. Start with MCD19A3 files from 2002-2019 that overlap with the PTA
    a. They contain 3 kernels per pixel that are used to back out BRDF from the RTLS equation for 8 bands. Choose the red band to characterize surface by visible reflectance.
2. Regrid data from step 1a to Albert's Equal Area projection over the PTA grid using the JPL provided bounding boxes. (This is the projection MAIA will use)
3. Retrieve the maximum BRF over 2002-2019 at each pixel for the subset of sun view geometry specified in
4.
5. ***Table 7***.
6. Use the maximum BRF data from every pixel in the PTA grid as input into K-Means clustering algorithm (discussed in **Chapter 3.4**). The sun view geometry information and BRDF over time is the feature vector for each pixel in the grid.
    a. Specify 16 clusters to map the data to.
    b. Specify land-water mask (from JPL MAIA AGP file)
        i. Assign water to 0 & coast to 1 manually before clustering (do not cluster this data)
    c. Run the algorithm which produces a classification for each land surface.
        i. If pixel had less than 5, 8-day periods of data from 2002-2019, mark as invalid (2) and assign it the nearest neighbor land surface ID
        ii. Order the classes by the mean BRDF of each of the 16 categories.
        iii. Class 3 is the darkest, class 18 is brightest

The reason max BRDF is chosen to represent the surface types is to record the brightest a surface can get under cloud free aerosol free conditions for a particular sun-view geometry. Anything brighter would imply a cloudy pixel or an aerosol filled pixel with $AOD_{0.47\mu m} \geq 1.5$ assuming the MAIAC snow, cloud and aerosol masks are accurate (Lyapustin et al., 2018). This is in line with the cloud conservative behavior by assuming the surface BRF is lower than that for clouds. The sun-view geometry space sampled is shown below in

Table 7. A composite graph of what is stored for input to the clustering algorithm is in **Figure 18** below. In this example it is the maximum BRDF observed for one pixel in the old Toronto PTA for each year from 2001 to 2016 for SZA equal to 45 degrees and DOY 185 of 365 (left). Then the data is combined by taking the maximum BRDF for all years (right). This datasets is produced for each of the pixels for every PTA, for 10 cos(SZA) and for 46 DOY bins.

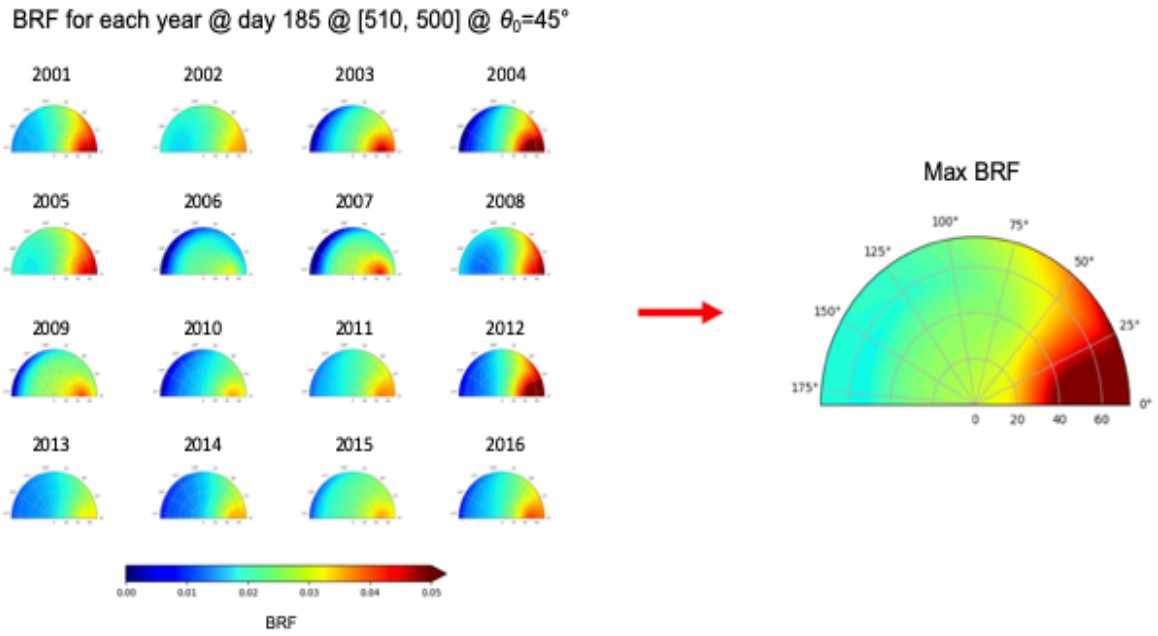

Figure 18: Example BRF polar plot for cos(SZA=45 degrees), DOY 185, at the pixel in row 510, column 500 in the fixed PTA grid. Figure produced by Dr. Yizhe Zhan.

Table 7: Sun View geometry values sampled for K-means cluster algorithm.

Quantity	Range	Interval	Unit
cos(SZA)	[0.05, 0.95]	0.10	[unitless]
VZA	[2.50, 72.5]	5	[degrees]
RAA	[7.50, 172.5]	15	[degrees]

In the next section the main surface ID algorithm and the analysis of its results are given using the workflow presented in this last section.

3.4. K-MEANS CLUSTERING LAND SURFACE ID (KLID)

This new database from the previous section is then used as input into the K-Means algorithm from the Sci-Kit Learn machine learning library in Python, implemented according to Arthur and Vassilvitskii (2007). This is an unsupervised clustering algorithm that assigns data points to randomly initiated cluster points in the feature space, and then iteratively moves the clusters until convergence. The feature space is the cos(SZA), VZA and RAA and the corresponding max BRDF for an 8 day interval. The input dataset is the BRDF at the sun-view geometry shown in

Table 7 for a single 8-day period. The other input is the number of clusters or labels to assign the data to. K-Means Clustering will initialize n random cluster centers in the feature space. We chose n to be 16 because it had the lowest misclassification rate (using MOD35 as the truth) for the LA PTA when it was used to run the MCM on all 18 years of Terra MODIS data as shown in **Figure 19**. It is worth noting though that the misclassification rate is a weak function of the number of land surface types used, but 16 was chosen because it resolved the surface features accurately. The algorithm then iteratively minimizes the distance between each randomly initiated cluster center and all the data points within the feature space by moving the cluster centers.

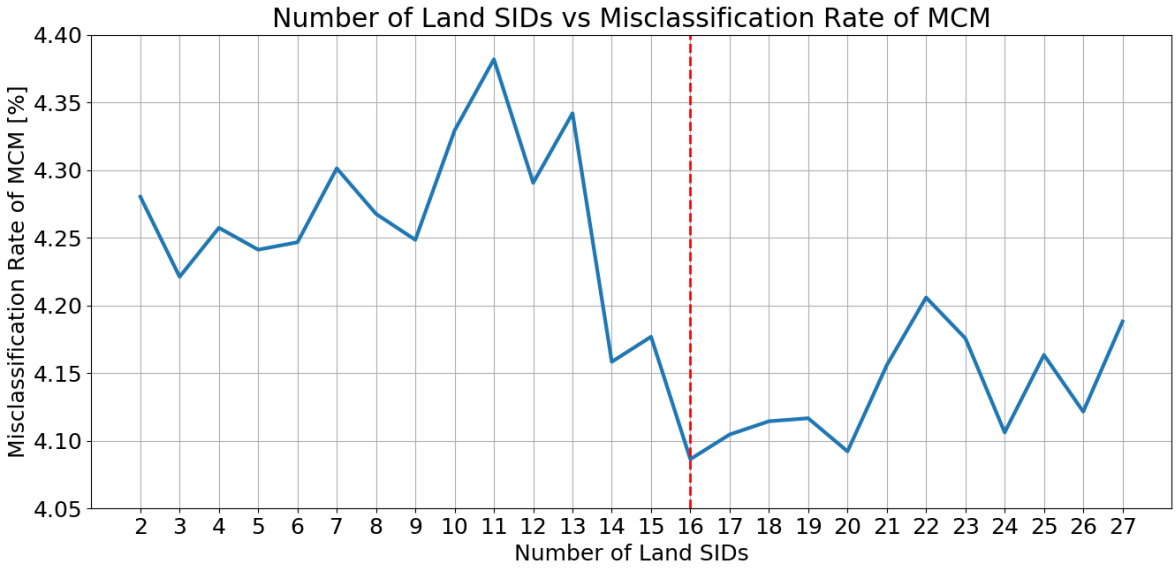

Figure 19: Misclassification rate in % of the MCM as the number of KLID surface types change, when MODIS is used as the true cloud mask. Red dashed line shows the global minimum of misclassification rate at 16 KLID surface types.

When the algorithm converges, each data point is assigned the cluster center closest to it. The result is that all the original data points are assigned to 16 clusters. Each group contains so many BRDFs and is now represented by the mean BRDF of the cluster. The categories are then reordered by the mean BRDF and assigned a value between 3 and 18. Where 3 is the darkest and 18 is the brightest land type. A land-water mask provided by JPL allows us to assign water to 0 and coast to 1. Points which did not have at least 5 years of input data to the 8-day interval or if any of the kernels from MCD19A3 were less than -1 were assigned the nearest neighbor land surface ID. The algorithm is run for every 8-day interval which gives 46 land surface ID maps for the year. In **Figure 20** 2 KLID maps are shown (category 3-18 is mapped to 0-15), one valid for DOY 185-192 and one valid for DOY 361-365. Because each cluster center is derived from one 8-day period of data (for years 2002-2019), each cluster, 0-15, will be a slightly different mean BRDF value from one DOY bin to another. This means that land surface category 4 on DOY 300 could have a

sufficiently different mean BRDF to category 4 on DOY 200 due to seasonal changes in albedo. Therefore, any one category is not equal to the same category in another 8-day period.

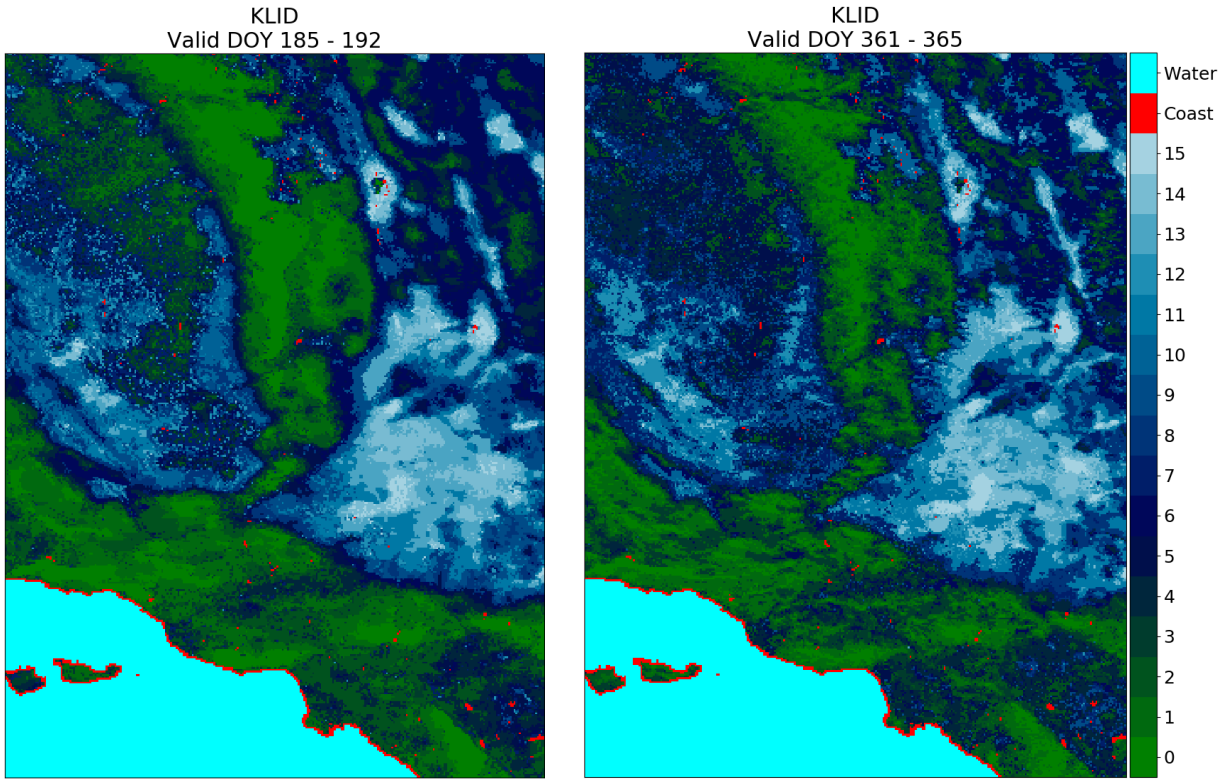

Figure 20: K-Means Surface ID maps valid at 2 different times of the year. 0 to 15 is derived from the algorithm. Coast and water categories come from a static land water mask provided by JPL. NOTE: The BRDF representing category X is not necessarily equal to the same X for another DOY bin, since they are derived from independent data.

Evaluation of KLID is done in two ways. 1) Compare it to a color image from a day within the DOY range to see if it is distinguishing the surfaces intuitively, and 2) see if the resulting cloud mask derived from KLID is any good. Currently KLID uses the nearest neighbor surface type to mask bad or missing data from the MAIAC dataset. In addition to this custom land surface ID, a water mask and snow-ice mask are needed, since precipitation and freezing can alter the surface properties. The threshold development uses the MOD35 snow-ice mask. The MAIA AGP file from JPL includes a static land-water mask that will be used for all processing. The MAIA science team

has decided to add the Global Multisensor Automated Snow and Ice Mapping System (GMASI) (Romanov, 2017) daily snow and ice mask to the cloud mask workflow since the MAIA mission is not producing its own snow-ice product. However, the current MCM threshold database is being developed with the MOD35 snow and ice product which is nothing more than a threshold of 0.4 on the normalized difference snow index (NDSI) and a cross check in warmer regions to agree with the National Snow and Ice Data Center (NSIDC) or National Oceanic and Atmospheric Administration (NOAA) snow-ice masks.

In the next chapter the threshold derivation using this surface ID, KLID, is shown.

CHAPTER 4 : PRE-LAUNCH THRESHOLD DEVELOPMENT

4.1. DATA: TERRA MODIS

The Terra MODIS instrument has a lot of characteristics similar and relevant to the MAIA mission. MODIS is a polar orbiting, wide swathed, multi-spectral imager, with a 1km resolution cloud mask with 4 confidence levels: cloud, uncertain clear, possibly clear, and clear. MAIA will also produce a 1km cloud mask in a similar orbit, with similar wavelengths and sun-view geometry. Terra launched in 1999 and at the time of this thesis it has been operational for 20 years – the longest NASA Earth Observing System (EOS) mission in history. Therefore, Terra MODIS provides the MCM with a long record of observations needed to produce robust thresholds. It should be noted that although MAIA is a multi-angle instrument like MISR, MISR lacks a lot of cloud masking hardware and VZA sampling that will be present on MAIA and is therefore not a good candidate to derive the thresholds with.

The thresholds are derived using Terra MODIS radiances, cloud mask, snow-ice mask, sun-view geometry and geolocation data from January 1st 2002 to December 31st 2019 (see **Table 8**). In **Table 9** MODIS bands 1, 2, 3, 4, 6 and 26 are used as proxy to MAIA bands 6, 9, 4, 5, 12 and 13, respectively. These will be the bands used for the cloud masking observables. Note that the band used for cirrus detection to derive the thresholds is 1.38 μm for MODIS while on MAIA it will actually be 1.88μm. Meaning that the thresholds would potentially need to be adjusted post-launch.

Table 8: MODIS products used in MCM threshold derivation.

MODIS Product File	Data Fields	Description
MOD021KM	EV_250_Aggr1km_RefSB Bands 1,2 EV_500_Aggr1km_RefSB Bands 3,4,6 EV_1KM_RefSB Band 26	Radiance [W/m^2/μm/sr]
MOD03	latitude/ longitude SolarZenith SolarAzimuth SensorZenith SensorAzimuth	Angles [degrees]
MOD35	Unobstructed_FOV_Quality_Flag	cloud, uncertain clear, possibly clear, clear
	Snow_Ice_Background_Flag	snow/ice, no snow/ice

Table 9: Cloud masking bands. Tan background (MODIS), blue background (MAIA), white background (MODIS and MAIA).

MODIS Band	1	2	3	4	6	26
Bandwidth nm	620-670	841-876	459-479	545-565	1628-1652	1360-1390
MAIA Band	6	9	4	5	12	13
Band Center nm	646	866	444	550	1610	1886
Description	Red	Water absorption	Blue	Green	Snow/ice absorption	Water vapor absorption

4.2. ALGORITHM FOR GENERATING THRESHOLDS

The MCM thresholds are derived using the same cloud conservative cumulative histogram method that MOD35 uses for its collection 6.1 visible reflectance test. The cloud conservative threshold is chosen such that 99% of the clear observations are classified correctly according to the MOD35 Unobstructed_FOV_Quality_Flag within a grouping of data known as scene type as seen in **Figure 21**. For the NDVI, NDSI and WI observables a slightly different method is used, which is described later.

Bin Name	Bin Range
cos(SZA)	0.9 - 1
VZA	40-45 deg
RAA	12-24 deg
Surface type	9
DOY	129-136
PTA	Los Angeles

Cloudy data	1308
Clear data	3162

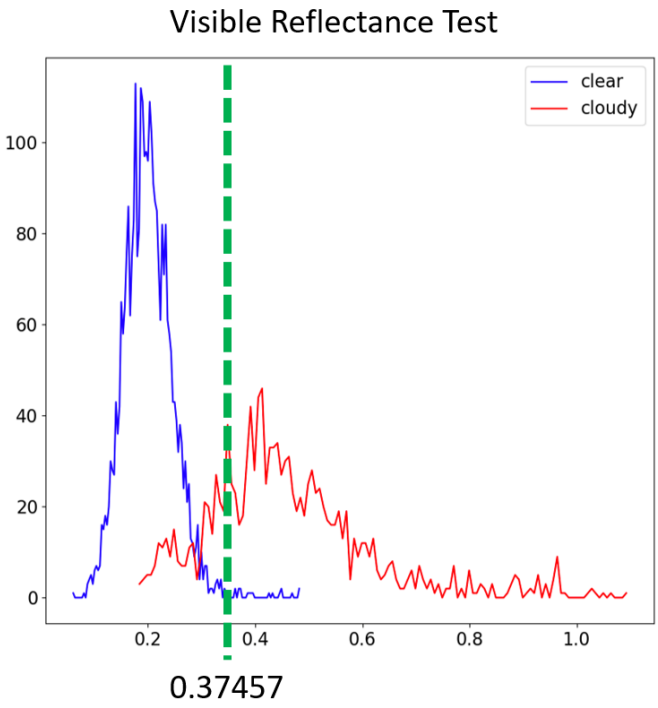

Figure 21: On the right is an example histogram of MODIS 0.64 μm BRF classified by its MOD35 cloud masking product used to create a threshold (green dashed line) for the data bins defined in the table on the left. Clear data is blue and cloudy data is red. The number of data points for each type is on the bottom left table.

A scene type is defined as an observation that is associated with a particular cos(SZA), VZA, RAA, scene ID and DOY bin (see **Table 11**). Scene ID is the KLID surface ID overlaid with the coast, water, sun-glint water and snow-ice masks. To create a threshold the scene type must contain 5000 data points belonging to clear pixels for Vis Ref, NIR Ref, Cirrus, SVI, and WI. For NDSI and NDVI the bin must have at least 5000 cloudy pixels. This is so there is enough data to capture the variability of the observable.

Table 10: Scene ID # along with meaning. Scene ID is subset of scene type.

Scene ID	Darkest Land	...	Brightest Land	Coast	No-Glint Water	Glint Water	snow-ice
#	0	...	15	16	17	18	19

Table 11: Every pixel is placed into these discrete bins to group like data.

Observable Level Parameters	# of bins	Bin Range	Bin Interval	Unit
cos(SZA)	10	[0,1]	0.1	[unitless]
VZA	14	[0,70]	5	[degrees]
RAA	12	[0,180]	15	[degrees]
Scene ID	15	[0,14]	1	[unitless]
DOY	46	[1,365]	8	[days]

This makes 10 cos(SZA)*12 RAA*14 VZA*20 Scene ID*46 DOY = 1,545,600 maximum possible scene types and therefore thresholds for each of the 11 PTAs and for each of the 7 observables. Because the observations are limited by the Terra orbit, MODIS camera, and the geography of each PTA, only a small subset of the sun-view geometry bins will be observed. This means many bins will not have a threshold derived for them. However, since MAIA will have a similar orbit to Terra and similar sun-view geometry to Terra MODIS this is not an anticipated problem. Nevertheless, for sun-view geometry bins with no thresholds, the nearest neighbor threshold is chosen while keeping the scene ID and DOY bin the same to preserve their meaning. Therefore, the nearest neighbor search is done by looking for the closest sun-view geometry bin that is populated with a threshold.

4.2.1. CODE STRUCTURE AND PERFORMANCE

The derivation of the thresholds was the main focus of my research for the master's degree. The code was developed to function in a highly parallelized environment that would not be possible without the resources of the University of Illinois at Urbana-Champaign's Department of

Atmospheric Sciences. Their Keeling computer cluster was used to store 5 Terra Bytes (TB) of data per PTA using anywhere from 1 to 46 to 200 cores at a time to carry out the processing and data analysis. What resulted was a modular and optimized suite of code to produce and analyze the thresholds with little prior knowledge of the code. What is presented is the logic to derive the thresholds for any PTA starting with a description of the main codes and the configuration file to customize it and ending with the batch script that runs all the code automatically in one command.

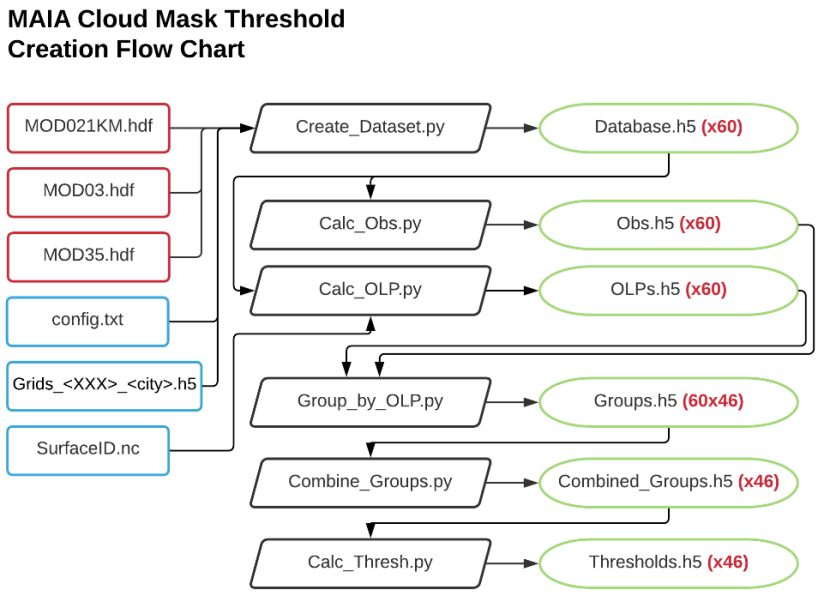

Figure 22: Flow chart showing the code and data used to produce the final threshold dataset. The red boxes on the left are the three Terra MODIS product files used (truncated to just the product name). The blue boxes are data files generated for the thresholds. Black rhombuses are individual scripts. Green ovals are datasets produced by the scripts. The red numbers indicate how many files were produced for that dataset. All file names slightly simplified for brevity.

Figure 22 shows the 6 scripts along with their input and output used to create the final threshold dataset. The *config.txt* file, shown in blue, is a human readable file that defines the paths to all the data needed to produce the thresholds. It is also editable so that when the code is run on a different machine the paths can be customized as the coder sees fit. The three red boxes on the left of **Figure 22** are the Terra MODIS 5-minute granules, described in **Table 8**. The

Grids_<XXX>_<city>.h5 files are defined for each PTA. The "XXX" is the country code, for example "USA" and the city could be "LosAngeles" to be *Grids_USA_LosAngeles.h5* for the LA PTA. It contains the exact longitudes and latitudes that define the 400 by 300 km MAIA PTA at 1 km resolution. Next the configuration file to run the code is described.

```
[current PTA]
#you can choose PTA from the PTAs section below
PTA: LosAngeles

[home]
home: /data/keeling/a/vllgsbr2/c/old_MAIA_Threshold_dev

[PTAs]
LosAngeles  : /PTAs/LosAngeles
Beijing     : /PTAs/Beijing
Barcelona   : /PTAs/Barcelona
AddisAbaba  : /PTAs/AddisAbaba
Delhi       : /PTAs/Delhi
TelAviv     : /PTAs/TelAviv
Rome        : /PTAs/Rome
Seoul       : /PTAs/Seoul
Taipei      : /PTAs/Taipei
Atlanta     : /PTAs/Atlanta
Boston      : /PTAs/Boston
Johannesburg: /PTAs/Johannesburg

[supporting directories]
MODXX             : PTA_MODIS_files/MOD
MODXX_paths_lists : PTA_MODIS_files/paths_lists
Database          : Database
Surface_IDs       : Surface_IDs
MCM_Input         : MCM_Input
obs               : thresh_dev/observables
OLP               : thresh_dev/OLP
group_intermediate: thresh_dev/group
combined_group    : thresh_dev/grouped_obs_and_CMs
thresh            : thresh_dev/thresholds
MCM_Output        : results/MCM_Output

[PTA lat/lon grid files]
LosAngeles  : /PTA_lat_lon_grids/Grids_USA_LosAngeles.h5
Beijing     : /PTA_lat_lon_grids/Grids_CHN_Beijing.h5
Barcelona   : /PTA_lat_lon_grids/Grids_ESP_Barcelona.h5
AddisAbaba  : /PTA_lat_lon_grids/Grids_ETH_AddisAbaba.h5
Delhi       : /PTA_lat_lon_grids/Grids_IND_Delhi.h5
TelAviv     : /PTA_lat_lon_grids/Grids_ISR_TelAviv.h5
Rome        : /PTA_lat_lon_grids/Grids_ITA_Rome.h5
Seoul       : /PTA_lat_lon_grids/Grids_KOR_Seoul.h5
Taipei      : /PTA_lat_lon_grids/Grids_TWN_Taipei.h5
Atlanta     : /PTA_lat_lon_grids/Grids_USA_Atlanta.h5
Boston      : /PTA_lat_lon_grids/Grids_USA_Boston.h5
Johannesburg: /PTA_lat_lon_grids/Grids_ZAF_Johannesburg.h5

[Target Area Integer]
LosAngeles  : 0
Beijing     : 1
Barcelona   : 2
AddisAbaba  : 3
Delhi       : 4
TelAviv     : 5
Rome        : 6
Seoul       : 7
Taipei      : 8
Atlanta     : 9
Boston      : 10
Johannesburg: 11
```

Figure 23: This figure shows the configuration file formatted in order to be queried by the Python library configparser. It contains the definitions of file paths and the PTAs that can be accessed by any scripts.

Figure 23 shows the contents of the configuration file, *config.txt,* that allows the paths of all the data to be edited and accessed from one file. This allows the user to work agnostically from the main code, which improves efficiency and avoids introducing new bugs into the main code. This is formatted to be accessed inside the main codes using the Python library, *configparser*. The headings in brackets define a category. The subheadings, without brackets, define variables where strings are stored. The strings can be changed to match the correct paths of the data for the working machine. In this figure, I show the working paths on the Keeling cluster where the thresholds were originally derived. The main purpose of this file is to change the working PTA, so the subheading *PTA* can be changed to any one of the 11 PTAs. As shown in line 3 it is set to the "LosAngeles" PTA. Now that the configuration file is defined the 6 scripts can be described.

Create_Dataset.py

This script takes the three Terra MODIS files corresponding to one 5-minute granule, truncates them to only what is necessary and reprojects them onto the current PTA. The rest of the data is not further used. The PTA is specified in *config.txt* and the longitudes and latitudes are defined in *Grids_USA_LosAngeles.h5* for the LA PTA for reprojection. The code is ran using 60 cores to reduce the run time from ~6.25 days to ~2.5 hours for one PTA. This results in 60 *Database.h5* files which contain a combined ~10,000 scenes from 2002 to 2019. Each scene is stored in a group named after the timestamp, <year><Julian day>.<UTC time> i.e. 2018365.1830. For each group, datasets are created for the radiances, reflectances, Earth-Sun distance, scale factors of radiance and reflectance, VZA, SZA, SAA, VAA, MOD35 cloud mask, latitude and longitude are stored in easy-to-read units ready for processing. The exact bands are listed in **Table 9**

Calc_Obs.py

This script reads in each *Database.h5* file and for each scene produces the 7 observables described in **Chapter 2.3**, the visible BRF, NIR BRF, Cirrus test, SVI, WI, NDVI and NDSI. The structure is identical to the 60 *Database.h5* files, meaning all the granules are in the same order, except only the observables are stored as datasets.

Calc_OLP.py

This script reads in each *Database.h5* file and the SurfaceID.nc file. For each scene, it produces the OLPs, or observable level parameters, as described in **Chapter 2.2**. The OLP defines the scene type of each pixel by its cos(SZA), VZA, RAA, DOY, surface type and PTA bin defined in **Table 3**. The structure is identical to the 60 *Database.h5* files, meaning all the granules are in the same order, except only the OLPs are stored.

Group_by_OLP.py

This is the most important script because it groups the data by its scene type which is defined by the OLP of each pixel. Along with the OLP of each pixel, the 7 observables and the cloud mask for that pixel are now known. That gives all the ingredients needed to calculate the thresholds. This script produces 60*46 or 2,760 files. This is because for each of the 60 *Obs.h5 and OLPs.h5* files, 46 files are written containing the unique scene types found in them. For example, all pixels belonging to DOY bin 0 are written to a file and so on to the 45^{th} bin for one corresponding set of OLPs.h5 and Obs.h5, so therefore 60*46 files total. Inside each file are groups defined for each unique scene type observed. Inside each group, each pixel that belongs to it is stored as that datapoint's 7 observables and its cloud mask as one row of data. Thus, a 2D array is created where

each row is one data point with 8 columns, the 7 observables plus the cloud mask. An example is shown in **Figure 24**. The group file in the green oval belongs to the LA PTA, with a timestamp in the first of the 60 OLP/Obs files and corresponds to DOY bin 0. Inside there can be *n* groups from 0 to 33,600. If a possible group (scene type) is not observed over the 18 years of Terra MODIS data, no group is recorded. Inside each group is a 2D array of 64-bit floats (the cloud mask is stored as a 64 bit float, but has values 0,1,2,3). The number of pixels that belong to a group is not known until the script ends.

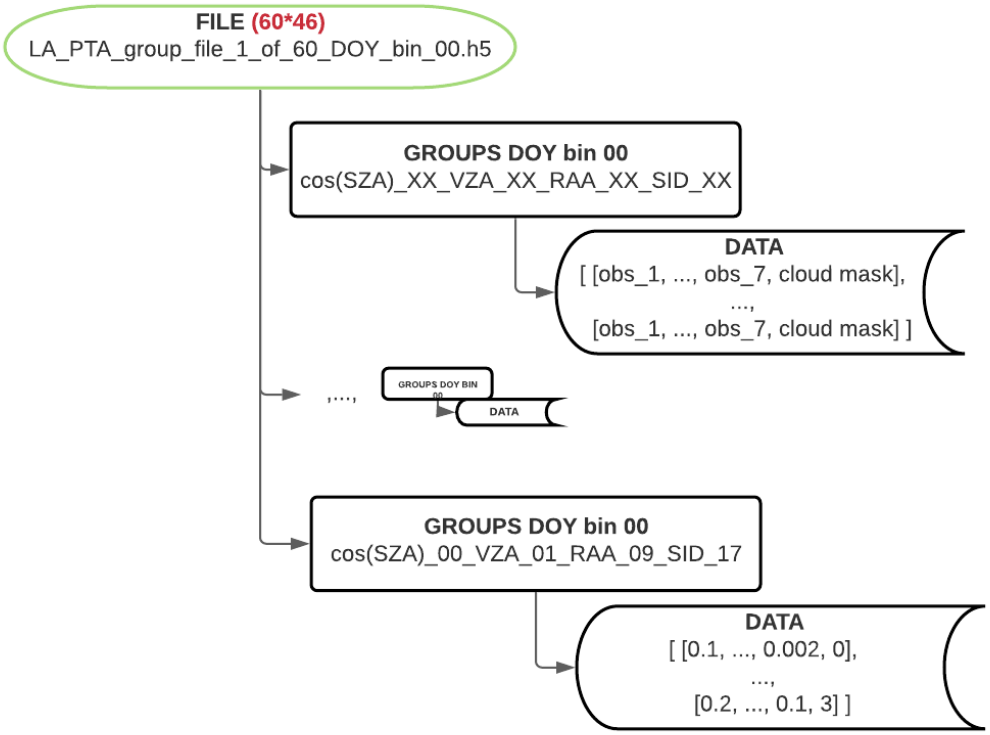

Figure 24: Structure of group files.

Therefore, the array has an unknown number of rows before the script is over. It could be 0 or 1 or 5,000,000 pixels that belong to that scene type. After this step the 2,760 files must be combined into 46 files. At the time of calculating the thresholds, all the data belonging to one group must be

read into memory, since the threshold is a percentile of the observables for one particular scene type (discussed further in **Chapters 4.2.2-4.2.4**). This is all done to minimize the memory and the runtime, since the code must search through all 18 years of data which cannot fit on the max 128 GB per node on the Keeling cluster. Moreover, this is to avoid redundant recalculations if a downstream script has bugs.

Combine_Groups.py

This script combines the 2,760 *Group.h5* files into 46 files, one for each DOY bin. This sets up the next script, *Calc_Thresh.py*, to be able to calculate the thresholds for each scene type. The reason to combine it into 46 files and not 1 file, is to run it parallelized on 46 cores. This dramatically improves the *Calc_Thresh.py* runtime from ~5 hours to ~7 minutes. The threshold derivation for each test is discussed in detail in **Chapters 4.2.2-4.2.4**.

```
1  #!/bin/bash
2  #SBATCH --job-name='make_thresh'
3  #SBATCH -p sesempi
4  #SBATCH --time=13:00:00
5  #SBATCH --mem-per-cpu=16gb
6  #SBATCH -n 60
7  #SBATCH --mail-type=FAIL
8  #SBATCH --mail-type=END
9  #SBATCH --mail-user=vllgsbr2@illinois.edu
10 #SBATCH --error='make_thresh.err'
11
12 ##goto working directory
13 cd /data/keeling/a/vllgsbr2/c/MAIA_thresh_dev/MAIA_CloudMask_Threshold_Development/scripts/
14
15 echo "create dataset"
16 mpirun -n 60 python MPI_create_dataset.py &
17
18 wait
19
20 echo "run observables"
21 mpirun -n 60 python calc_observables.py &
22
23 wait
24
25 num_SIDs=16
26 echo "run observable level parameter"
27 mpirun -n 60 python calc_OLP.py $num_SIDs &
28
29 wait
30
31 echo "group data by OLP and combine with obs"
32 for i in {0..45}
33 do
34     echo $i
35     mpirun -n 60 python group_data_by_OLP.py $i &
36     wait
37 done
38
39 echo "combine grouped files by DOY"
40 mpirun -n 46 python combine_grouped_files.py &
41 wait
42
43 echo  "run thresholds"
44 mpirun -n 46 python calc_threshold.py $num_SIDs
45
46 exit 0
```

Figure 25: Batch script for the SLURM job manager used on the Keeling cluster. This shows how all codes are run in one central script using multiple cores.

Figure 25 shows the batch script to submit the codes to the SLURM job manager. This allows the 6 scripts to be run from one central location after debugging. Using this method, the wall time of 13 hours (line 4), 16 GB of memory per core (line 5), 60 cores (line 6), a notification system (line 9) and an error file (line 10) can be defined to run the scripts. Then the working

directory is defined (line 13) where all the code is stored. Afterwards, each code is run with the number of cores it requires using the *mpi4py* Python library to pass it to the scripts. After every script there is an "&" and in the line below the key word "wait", which tells the SLURM job manager to run the codes linearly, so when one script's final data is written onto the disk, then the next script can access those files after they have been fully populated. The first four scripts on lines 16, 21, 27 and 35 get 60 cores and 16 GB of memory per core. This was chosen to split the processing work 60 ways, which results in 60 separate files per script. The 60 files from *MPI_Create_Dataset.py* feed into *calc_observables.py*, which produces 60 files that feed into *calc_OLP.py* and so on into *group_data_by_OLP.py*. This allows data to be written to disk without errors from multiple cores accessing the same file. However, these need to be combined to produce the thresholds, so on line 40, *combine_grouped_files.py* is ran to produce one file for each DOY bin, 46 in total. Finally, *calc_threshold.py* is ran to create 46 separate threshold files, one for each DOY bin. In the main MCM code from **Chapter 2** the correct threshold file is chosen automatically based on the timestamp of the scene given.

All the code was run on the Keeling cluster using the Intel(R) Xeon(R) CPU model E5-2660 v3, clocked at 2.6 GHz with 1 thread per core and 10 cores per socket. Each node has 20 cores and 128 GB of memory. The allocation of cores and memory used is shown in **Figure 25**.

4.2.2. VISIBLE REFLECTANCE/ NIR REFLECTANCE/ CIRRUS TEST/ SVI THRESHOLD DERIVATION

These observables have the same threshold derivation since they all correlate positively with clouds. In other words, higher values of these observables are associated with higher cloud confidence. Lower values are associated with higher clear confidence. The more reflective a pixel is in the NIR or visible channel, the more textured it is in a 3x3 pixel window, the more reflective

it is in the water vapor absorption channel the more likely it is that that pixel is cloudy. Therefore, exceeding the threshold should flag the pixel as cloudy. The threshold is chosen to be the value, such that 99% of the clear pixels for a group (scene type) are classified correctly according to the cloud confident MOD35 classification (**Figure 21**).

In **Figure 26**, the pseudo code to calculate the threshold for these tests is shown. First it checks if there are clear data points for "obs_X" and for scene type "xyz". If there are clear data points in that group, the threshold is chosen to be the 99th percentile of those values using the nanpercentile function from the Python library, *numpy*.

```
#VIS/NIR/SVI/Cirrus
if there_are_clear_data_points:
    #one of VIS/NIR/SVI/Cirrus
    clear_obs_x_bin_xyz = clear_obs_x_bin_xyz['choose_one_observable_in_a_loop']
    #get 99th percentile of observable x from all data belonging to bin xyz
    threshold_for_bin_xyz = np.nanpercentile(clear_obs_x_bin_xyz, 99)
```

Figure 26: Pseudo code of how to calculate threshold and where to store it in threshold array. bin_xyz is the scene type, obs_x is 1 of 7 observables.

4.2.3. WHITENESS INDEX (WI) THRESHOLD DERIVATION

The whiteness threshold is derived similarly to the above, but WI values are negatively correlated to cloudiness. In other words, a WI of 0 is perfectly white and higher values are less white. Since clouds are white in the visible spectrum, they are associated with values close to 0. So instead of exceeding a threshold to be cloudy, the pixel must fall below the threshold to be cloudy. The threshold is chosen to be the value, such that 99% of the clear pixels in a group (scene type) are classified correctly according to the cloud confident MOD35 classification.

In **Figure 27**, the pseudo code to calculate the threshold for this test is shown. First it checks if there are clear data points for "obs_X" and for scene type "xyz". If there are clear data points in that group, the threshold is chosen to be the 1st percentile of those values using the nanpercentile function from the Python library, *numpy*.

```
#WI
if there_are_clear_data_points:
    #get 1st percentile WI value from all data belonging to bin xyz
    threshold_for_bin_xyz =  np.nanpercentile(clear_obs_WI_bin_xyz, 1)
```

Figure 27: Pseudo code of how to calculate threshold and where to store it in threshold array. bin_xyz is the scene type, obs_x is 1 of 7 observables.

4.2.4. NORMALIZED DIFFERENCE VEGETATION AND SNOW INDICES (NDxI) THRESHOLD DERIVATION

These thresholds are derived differently. Values approaching zero from the negative and positive sides are associated with cloudy pixels, and values far from zero are associated with water, snow and vegetation. Therefore, we need to create a range centered around a threshold, and any pixel with NDxI that falls within that range is called cloudy. To understand why the threshold is calculated the way it is, the metric distance to threshold or DTT, must be understood. DTT is the Euclidean distance between the threshold and the observable value, normalized by the threshold. Because NDxI can be negative, the absolute value is taken. This gives the equation

$$(T - |NDxI|) / T \qquad \textbf{eq. (11)}$$

The result is that for a given threshold, DTT creates a window or range around the threshold for which those NDxI values correspond to cloudy pixels. In other words, any NDxI value larger in magnitude than the threshold resulting in DTT < 0, is outside this range and therefore from a clear pixel, otherwise it falls within the range and is therefore from a cloudy pixel, corresponding to DTT >= 0.

The threshold is chosen as the value corresponding to the most populated bin in an NDxI histogram of cloudy pixels, which is the peak of the histogram or mode of NDxI for cloudy pixels. The NDxI values are binned from -1 to 1 over 128 bins. A negative DTT then means the NDxI value is outside this range and therefore clear. Note that this method assumes the cloudy signal is symmetric about the threshold. This is rarely the case, however. The real range is always slightly

lopsided and is necessarily a limitation of this technique. Despite this, the results are still good. A possible improvement would be to independently choose two thresholds, one to bound each side of the range, instead of relying on DTT which is symmetric.

In **Figure 28**, the pseudo code to calculate the threshold for these tests is shown. First it checks if there are cloudy data points for "obs_X" and for scene type "xyz". If there are cloudy data points in that group, a histogram of the data is built using the *histogram* function from the Python library, *numpy*. It is binned from -1 to 1 with 128 bins. Then the mode of the histogram is chosen as the threshold by finding the bin that has the maximum number of samples. Since the mode is not unique to one bin in a histogram, the minimum of all the modes is chosen as the threshold. This is because clouds tend to lower the NDVI, so the lowest mode is the most cloud conservative.

```
#NDxI
if there_are_cloudy_data_points:
    hist, bin_edges       = np.histogram(cloudy_obs_NDxI_bin_xyz, bins=128, range=(-1,1))
    #threshold is bin with most cloudy pixels
    threshold_for_bin_xyz = bin_edges[1:][hist==hist.max()].min()
```

Figure 28: Pseudo code of how to calculate threshold and where to store it in threshold array. bin_xyz is the scene type, obs_x is 1 of 7 observables.

CHAPTER 5 : PRE-LAUNCH VERIFICATION OF MCM AND THRESHOLDS

5.1. THRESHOLD ANALYSIS AND ALGORITHM PERFORMANCE

In **Chapter 4** the thresholds are derived for the MCM. Analyzing the values of these thresholds gives insight into how the tests characterize clouds and the background for each scene type. The analysis that follows is for the LA PTA using the Terra MODIS dataset from 2002 to 2019. All accuracy referred to are training accuracies since the Terra MODIS data used to evaluate and to derive the thresholds are the same. The MOD35 cloud mask from Terra MODIS is assumed to be the truth where confident cloudy is "cloud" and uncertain clear, probably clear and clear are all "not cloud".

The histograms of the thresholds for each observable are shown in **Figure 29** for the LA PTA. These histograms combine all surface ID, DOY, and sun-view geometries, and are derived from MODIS-Terra data collected between 2002 to 2019. The y-axis is plotted on a log scale to show the less full bins. The histograms show the wide range and frequency of thresholds derived to closely match the Terra MODIS MOD35 confident-cloud cloud mask. The range of thresholds come about from the binning used in **Table 3**. This binning was chosen to create independent scene types such that all the data in one scene type share similar physical attributes, as to derive a threshold that can best separate both classes of observations, cloud and not cloud. The WI histogram is bi-modal where most of the thresholds are near zero (the first hump), suggesting a split in the data where one group of scene types (e.g., vegetated surfaces) is much less white than the others (e.g., desert surfaces) given the observed clouds. The second hump could correspond to non-sun-glint water which is very black. This second hump can be seen in isolation in **Figure 30** for the WI water boxplot. NDVI has most of its thresholds between 0 and 0.4. NDVI greater than 0.4 starts to become vegetated land and below zero is characteristic of water, while thick clouds

will tend the NDVI to just above zero (Zhu and Woodcock, 2012). The NDSI thresholds are all above 0.4. This is because the snow mask used to derive the MCM thresholds uses the condition of at least 0.4 NDSI for snow (Ackerman et al., 2010). The histogram suggests the observed clouds over snow according to the MCM occur at NDSI values between 0.4 and 0.7. The visible BRF threshold histogram peaks around 0.2 and extends to BRFs over 1. The NIR BRF thresholds are only applied over non-sun-glint water and thus tend to be a magnitude lower in value than the visible BRF thresholds, since water is very dark at 0.86 μm BRF but clouds are still relatively bright (Ackerman et al., 1998). The higher thresholds observed are due to shallow waters and sandy beaches classified as water in the land-water mask. The sun-glint mask tends to do a good job of masking out the glitter regions and is not considered as significant contributing factor to the high thresholds. The SVI thresholds peak around 0.05. The higher SVI values are due to the observable also being applied over land which has much higher texture than the ocean at 1km resolution. The cirrus threshold histogram has a global maximum between 0.025 and 0.05 then falls quickly with increasing BRF.

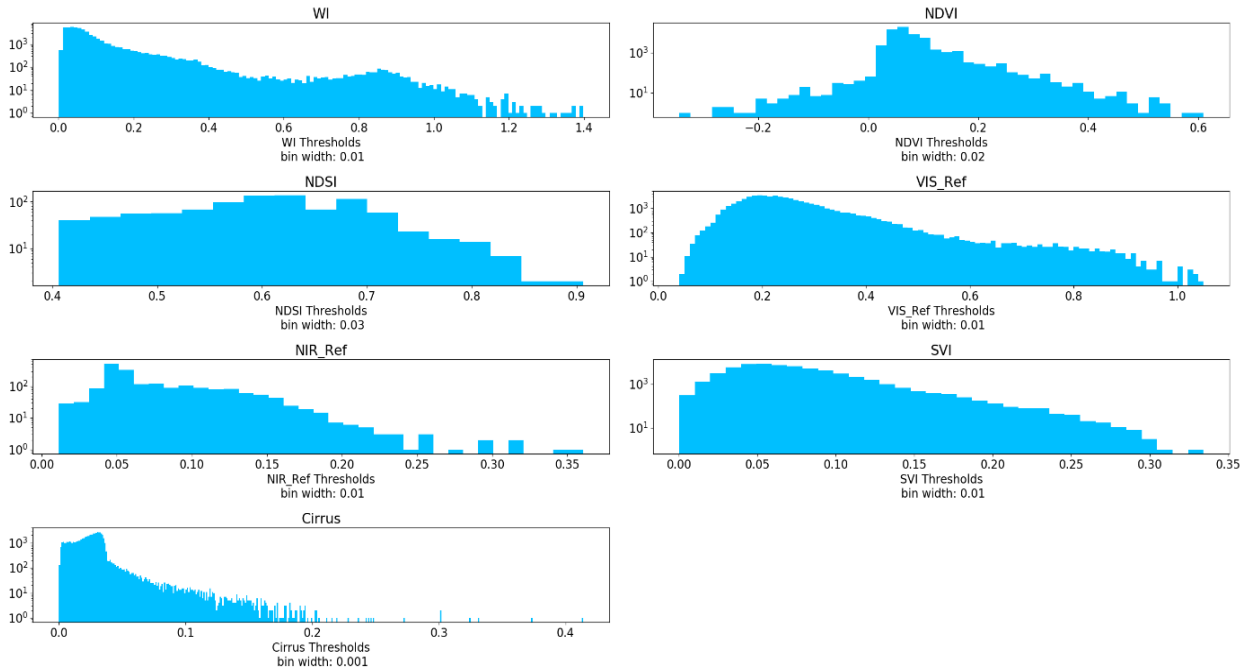

Figure 29: Histograms of the complete threshold dataset for each of the 7 observables for all surface ID, DOY and sun-view geometry bins. The x-axis is on a linear scale. The y-axis is on a log scale and denotes the number of thresholds in that bin.

The box and whisker plots of the thresholds for each observable (**Figure 30**) show how the distribution of thresholds changes as a function of surface ID. This gives insight into how surface types contribute to the thresholds distributions and if each one gives a unique advantage to the MCM. The distributions of thresholds are only shown when the observable is applied to that surface ID. The thresholds are valid for the LA PTA, derived from years 2002 to 2019. The lines on the end of the whiskers are the range excluding outliers, the box is the 1st and 3rd interquartile ranges and the orange line in the middle is the median. The red line overlayed on the box and whisker plots show the percent change of the mean threshold from surface type n-1 to surface type n, from n=1 to n=16 which is the coast category. This is helpful to see how strong of a function the thresholds are of the surface type as the surface gets brighter. Water and NDSI are only used over one surface type, so no meaningful threshold analysis can be done with respect to changing surface type.

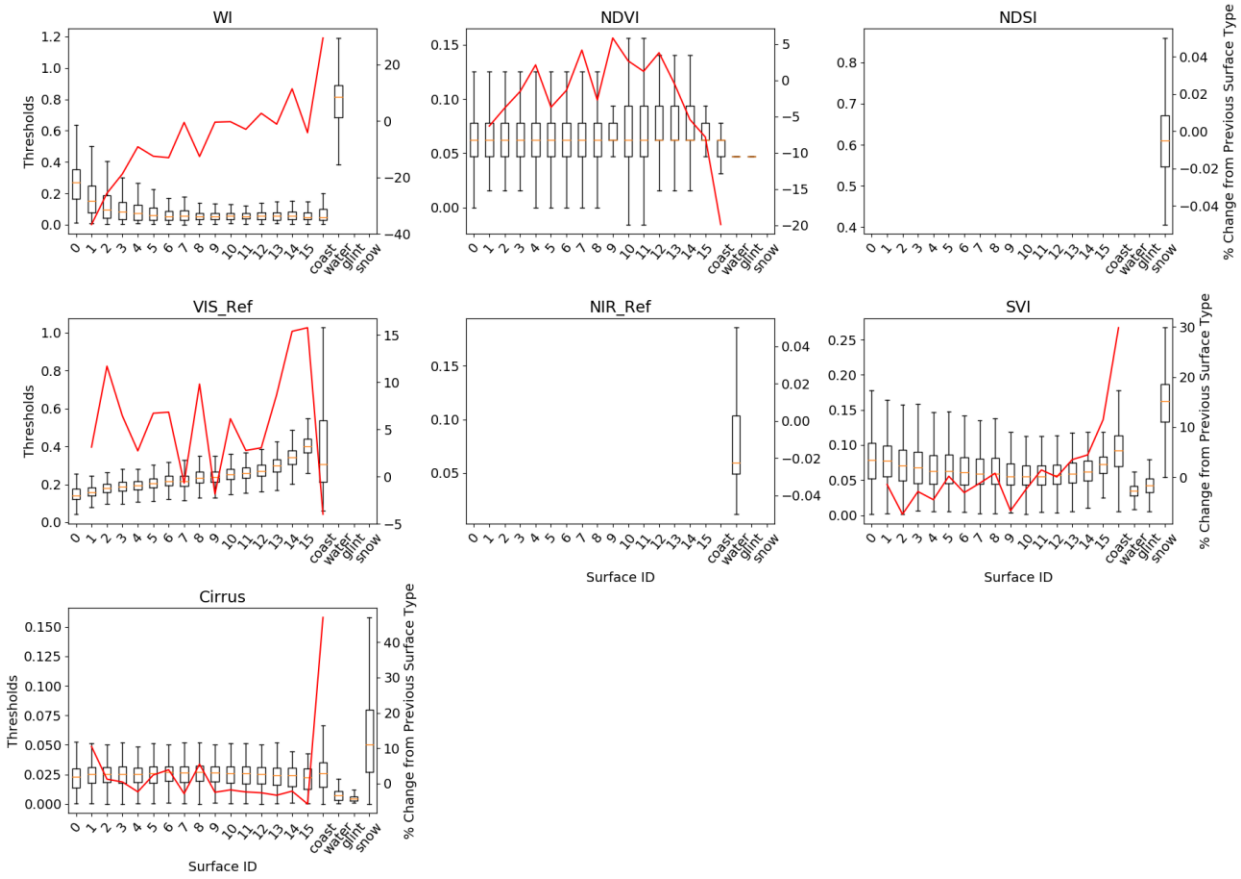

Figure 30: Box and whisker plot to show the distribution of thresholds for each surface type and for each observable independently. The x-axis is the surface type from KLID, 0 to 15, and the rest are from ancillary masks. The left y-axis shows the threshold values and the right y-axis (red line) is the percent change of the mean threshold from the previous surface type valid from 0 to coast. Box and whisker plots are only produced when that observable is applied over that surface type.

For the WI, a perfectly white pixel has the value of 0, and the value becomes more positive for less white observations. The WI thresholds here decrease with the KLID surface types implying that the brighter surfaces like the deserts in the LA PTA are the whitest and therefore need the lowest WI thresholds to decern clouds from the background. This means cloud detection over the brighter surfaces is a more difficult than over darker surfaces for the WI since it will necessarily miss clouds with higher WI values. Non-sun-glint water is the blackest surface type the MCM recognizes and therefore has the largest WI thresholds. This indicates over this surface the WI has

highest contrast between clouds and the background, able to find thinner clouds than possible over bright desert.

For the NDVI the threshold distributions vary relatively weakly with surface type around 0.05. However, some surface types show more variability in thresholds than others. Notably for surface type 9, 15, coast, water and sun-glint water.

For the visible reflectance test the thresholds increase with KLID surface types as expected since they are ordered by BRF from darkest to brightest. The largest threshold range is for the coast surface type which is likely due to the inherit errors in the land-water mask. This can be seen by inspection, where some coast pixels contain deep ocean, some shallow ocean, some sandy beaches, and some dark land. This suggests it maybe more useful to let KLID decide the surface type of the pixels belonging to the coast type instead of a static land-water mask.

For SVI the threshold distributions first decrease towards zero with surface type but then increase. Vegetated land tends to occur on flat terrain and vary smoothly in the LA PTA, whereas the brighter desert regions contain more boundaries and rough mountainous terrain. Coast has the highest snow-free land thresholds of SVI which is likely due to the coast type being a surface boundary by definition. Snow has the largest over all SVI thresholds which is consistent with the mountainous region that snow is found in for the LA PTA. The snow surface on mountains also have high SVI values because the perimeter of the snow fall is very irregular. Water has the lowest SVI thresholds followed by sun-glint water which is expected. At 1km resolution most waves cannot be resolved, but sun-glint causes texture associated with how sun-glint becomes visible. It also makes the banding in the MODIS images more obvious which can lead to higher texture values.

The cirrus thresholds vary little with the KLID surface types, 0-15. Often the lower atmosphere has enough water vapor to obscure the surface from view (Gao et al., 1993) and make the surface type irrelevant. Despite this, the mean thresholds do change about 5% to 10% across the surface types indicating something about the detectability of clouds with respect to the KLID surface types, which may indicate spatial variations in the amount of water vapor that is closely tied to the surface type, such as dry deserts. The mean coast threshold is over 40% greater than the mean thresholds for surface type 15. No reason is apparent from the time of this thesis. The two water types have the lowest cirrus thresholds which is consistent with water's high absorption in the NIR wavelengths (Jedlovec, 2009). This implies clouds above the moist layer have higher contrast over water than over land, particularly for a dry atmosphere.

The half polar plots (**Figure 31**) show the training accuracy of the MCM (as compared to Terra MODIS) as a function of the sun-view geometry for the LA PTA from 2002 to 2019. There are 10 polar plots, one for each SZA bin. The azimuthal axis is the RAA with symmetry about the principal plane with the backscatter direction defined at 180 degrees and the forward scatter direction at 0 degrees. The zenith or radial axis is the VZA. Every grid cell represents a unique combination of SZA, RAA and VZA. The empty cells are where not enough data were observed belonging to that bin from Terra MODIS to produce a threshold (note that no nearest neighbor thresholds were used to occupy the empty cells, since no data from MODIS has sampled that sun-view geometry over the LA PTA). The lowest accuracies are seen in the bins adjacent to empty cells in the direction of increasing VZA. This may suggest a sampling issue where there are not enough high VZA observations able to inform the threshold calculation before the cutoff of 5,000 required to derive a threshold. The highest fraction of grid cells is filled for low SZA which is consistent with Terra's 10:30am equator crossing time where the sun is near its max height, and

with the LA PTA's latitude around 34 degrees North. As the SZA decreases the forward scatter direction performs better than the backscatter direction. In **Figure 32** the exact same plot is shown, but for the number of samples used to compute the accuracies in the previous figure. The many low accuracy bins coincide with low sample numbers below 10^8 observations. This suggests the MCM has difficulty choosing a threshold for these bins due to sampling. However, sample number alone cannot explain why the MCM fails in certain bins, and this is not easily verified by image inspection since the sun-view geometry bins vary rapidly within a MODIS granule. Further analysis is required to determine the cloud and surface properties that would lead to low performance under these sun-view geometries.

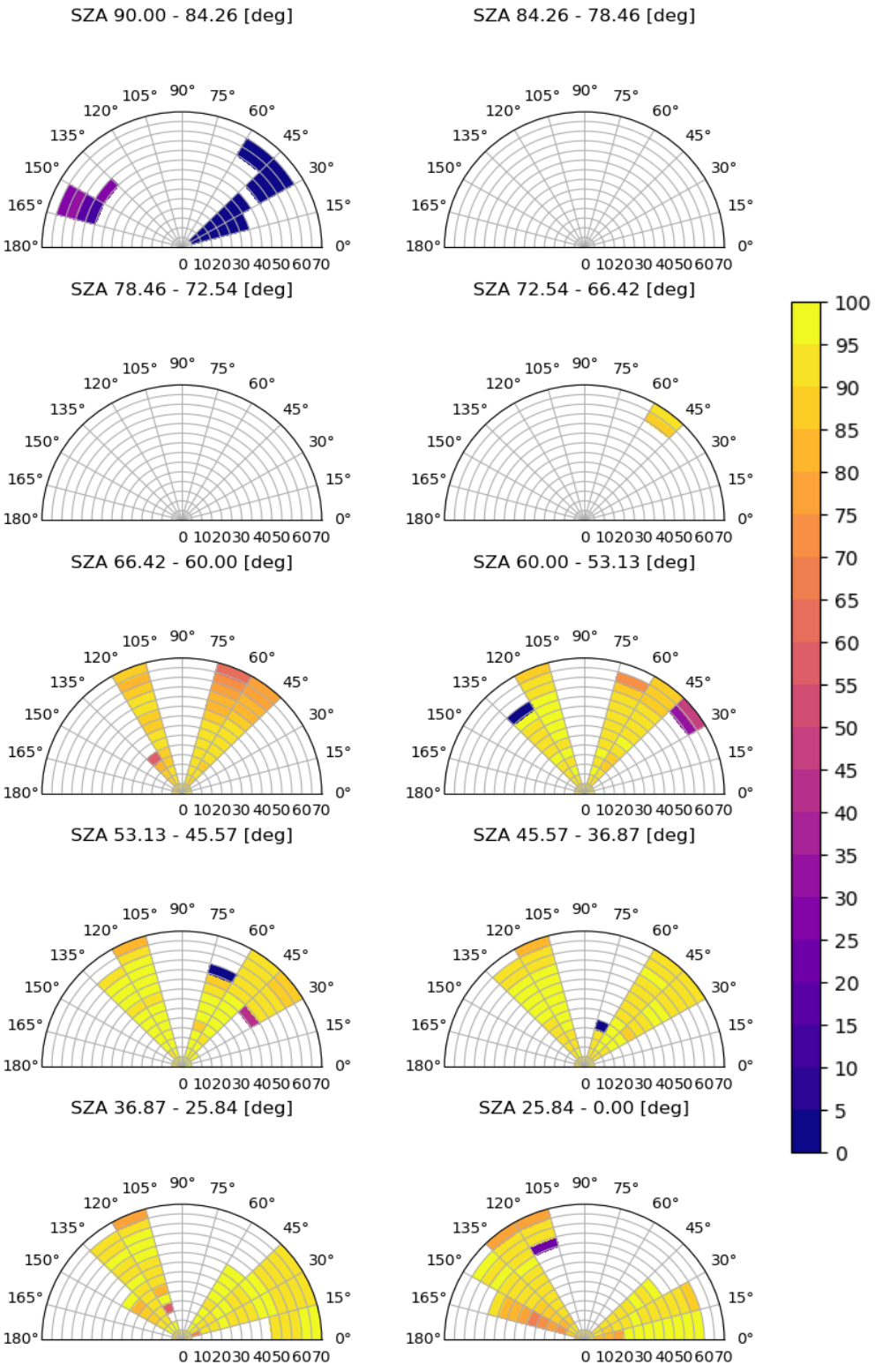

Figure 31: Training accuracies plotted for each sun-view geometry bin observed in the dataset. Each plot is a SZA bin (left to right top to bottom the bins go from 9 to 0). The azimuthal axis is the RAA. The zenith or radius axis is the VZA. Note the VZA bins go every 5 degrees on the grid lines, but only every 10 degrees is labelled to make the plot readable.

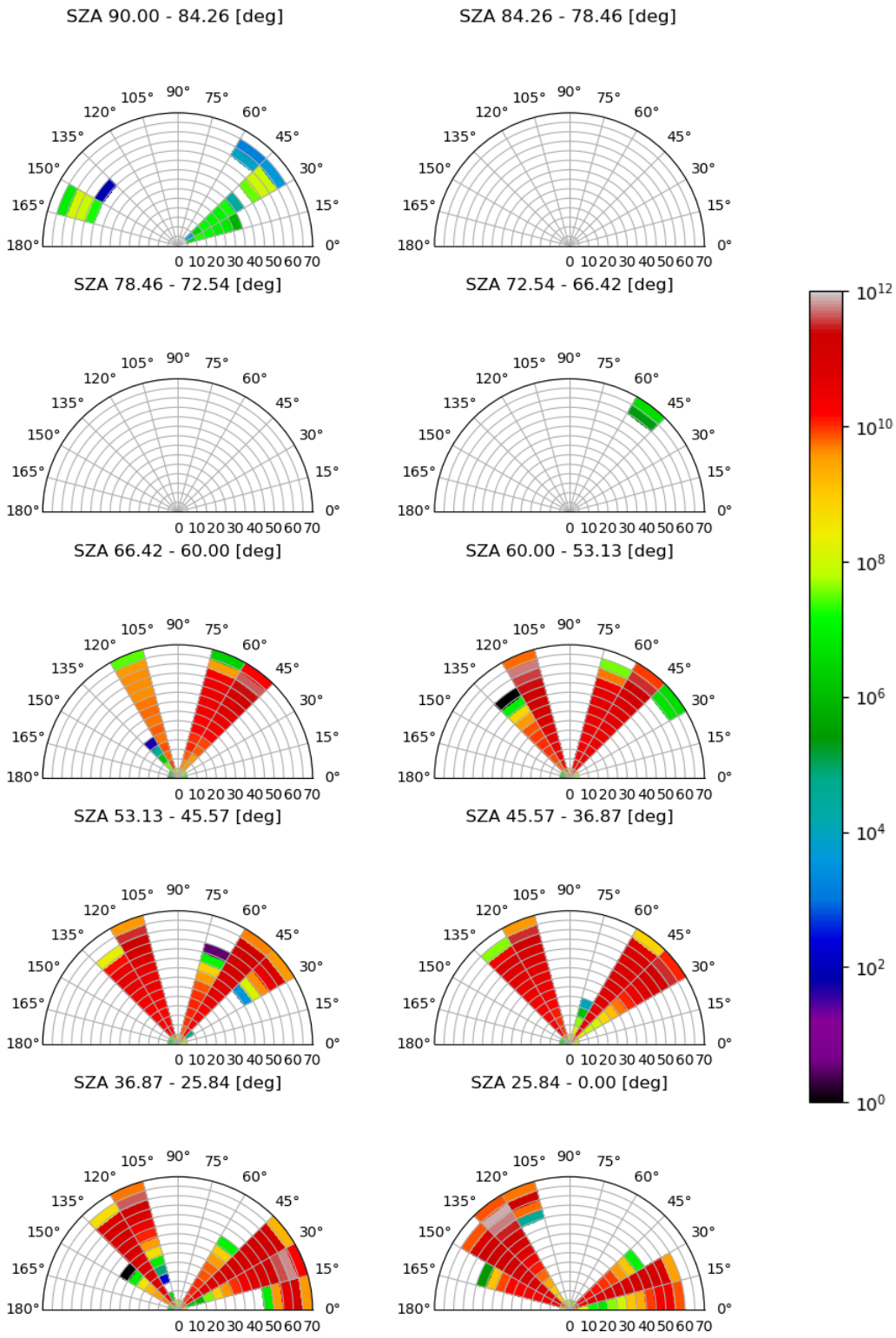

Figure 32: Number of samples plotted for each sun-view geometry bin observed in the dataset. Plot is in log base 10 scale. Each plot is a SZA bin (left to right top to bottom the bins go from 9 to 0). The azimuthal axis is the RAA. The zenith or radius axis is the VZA. Note the VZA bins go every 5 degrees on the grid lines, but only every 10 degrees is labelled to make the plot readable.

There are 16 KLID surface types plus coast, water, sun-glint water and snow-ice. Two important questions are 1) how is the MCM doing over land, water and snow, and 2) are the KLID surface types performing as expected. The pink line in **Figure 33** shows the accuracy of the MCM (as compared to Terra MODIS, MOD35, for the LA PTA from 2002 to 2019) as a function of surface type and the cyan line shows the number of samples for the surface type in a log scale. Most notable is that the log scale of the number of samples line correlates well with the accuracy line implying low performing surface types could benefit from an increase in observations. In spite of this, the accuracies still reflect the relative difficulty of cloud detection over each surface type. Non-sun-glint water has 95% accuracy, snow performs the worst with 42% accuracy, and the accuracy decreases from ~97% to ~81% as the surface types get brighter from 0 to 15. The coastline often appears cloudy in clear scenes and is the worst performing after snow at 80%.

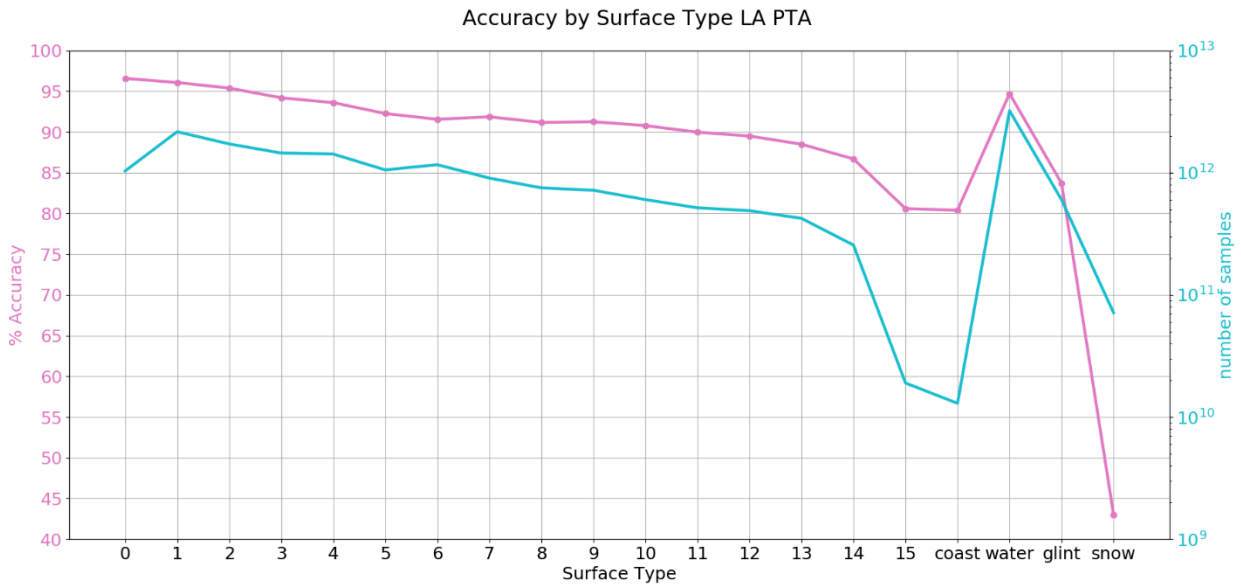

Figure 33: Accuracy of the MCM over the LA PTA from 2002 to 2019 by the surface type. The left axis in pink shows the accuracy in percent and the right axis shows the number of samples in the surface type on a log scale.

Figure 34 shows the accuracy at a fixed location in the grid for the LA PTA by the four seasons from 2002 to 2019 (the accuracy is computed relative to the number of samples in a fixed location). The seasons are defined by DJF (Winter), MAM (Spring), JJA (Summer) and SON (Fall) which are the months December, January, February, March, April, May, June, July, August, September, October, and November. For reference, **Figure 35** shows a KLID map valid for the DJF period. Although the map varies every 8 days, the basic patterns are the same. The highest accuracies (above 95%) observed in all panels coincide with the green surface types from category 0 to 3. These are the 4 darkest land surface types and give the most contrast to cloud cover. The red areas in **Figure 35** corresponding to the coast and have some of the poorest performance, near 40% to 60%. The coast is bright and contains a mixture of land and water types that use the same thresholds which makes it an inherently difficult region for the MCM. The Sierra Nevada mountains have KLID values greater than 7 and also exhibit low accuracies around 60% to 70%. These two figures show well how the brightness of the surface type correlates negatively with the performance of the MCM. This suggests and is later supported by the confusion matrix (**Table 12**) that the dominant error is due to false cloudy flags caused by bright surfaces. The seasonal variation in performance is less dramatic. The dominant signals from the coast, the mountains, and the bright surfaces still hold up, however, the Sierra Nevada mountain area performs worst in the summer (JJA). JJA has the lowest snow amount of all the season and is the driest and hottest time of the year for this region. Further investigation would be needed to find out what in the MCM or observations caused the drop in performance.

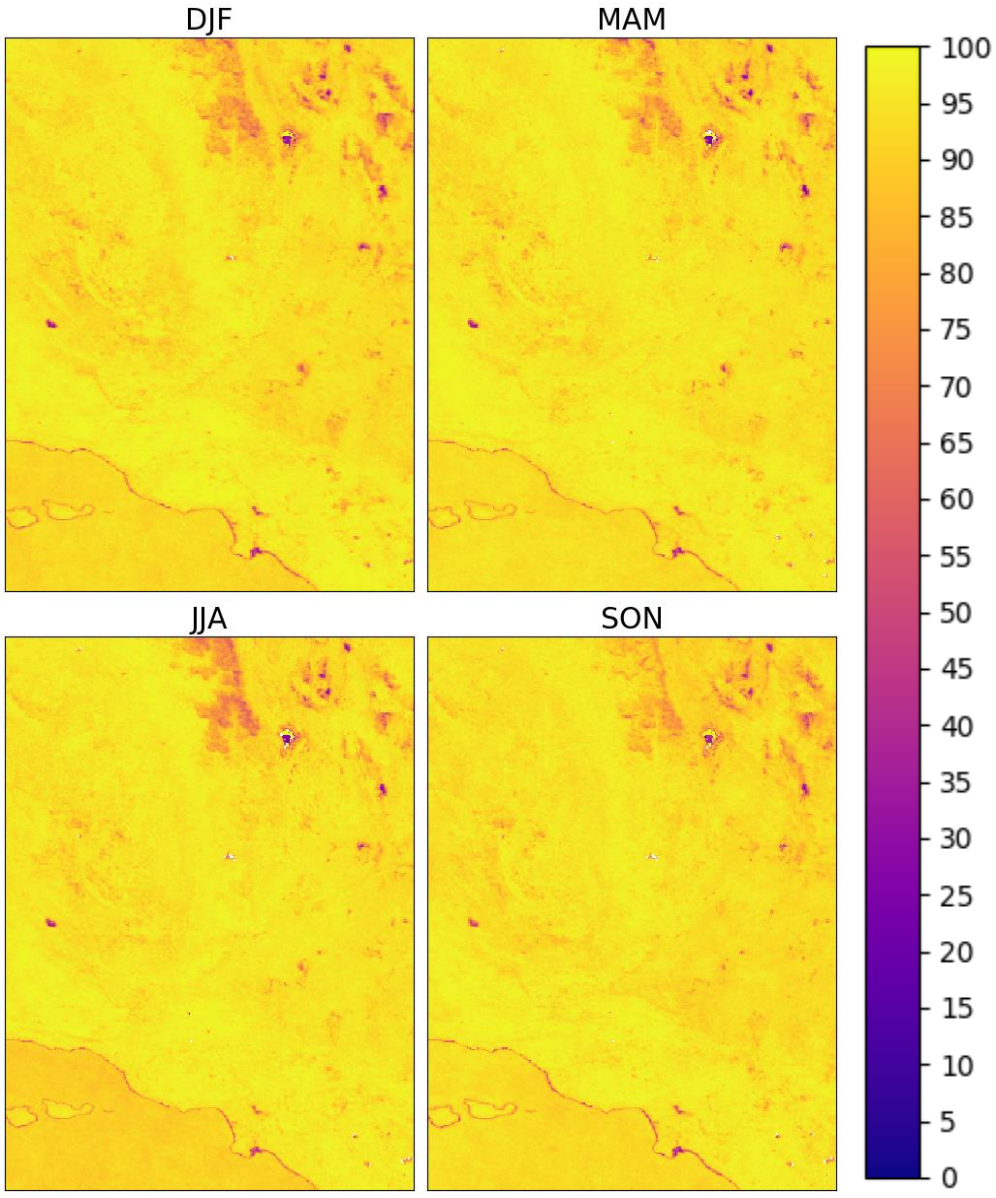

Figure 34: Training accuracy of the MCM of the LA PTA by season as compared to MOD35 confident cloudy class. Accuracy in each pixel is valid for that fixed location from 2002 to 2019. DJF, MAM, JJA and SON are the months December, January, February, March, April, May, June, July, August, September, October, November.

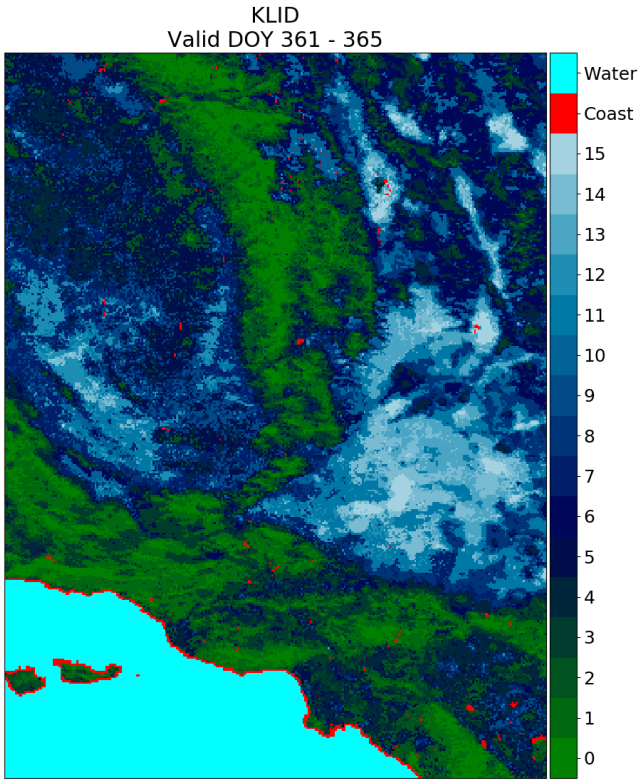

Figure 35: The KLID map of the LA PTA valid for DOY 361-365 which are the last 6 days of December.

Figure 36 shows the accuracy of the MCM and the number of samples as a function of the 46 DOY bins for the LA PTA from 2002 to 2019. The DOY on the x-axis represents that Julian day and the previous 7 days of data. The performance is good throughout the year and both lines are somewhat in phase. The DOY 24, 48, 88, 232, 288, and 344 accuracy local minima have coincident local minima with the sample number suggesting these DOY bins could improve accuracy if they had more observations, although that does not seem to explain all the minima in accuracy, so further investigation is required. And lastly, the worst performing DOY bins are later in the year and get worst as the year ends.

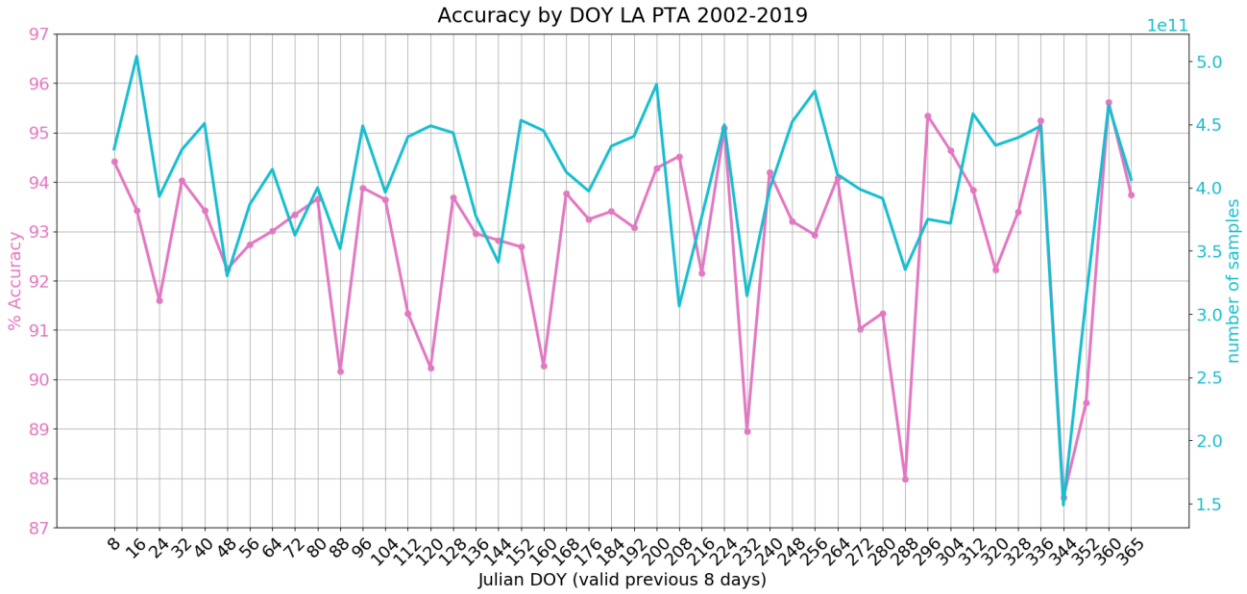

Figure 36: Training accuracy as compared to MOD35 confident cloudy class over the LA PTA from 2002 to 2019 by Julian DOY. The pink line shows the accuracy, and the cyan line shows the number of samples used to calculate the accuracy. The sample number is scaled by 10 to the power of 11.

The confusion matrix for the MCM (**Table 12**) shows for how many pixels the MCM and MOD35 cloud-confident cloud mask agreed and disagreed on for the cloudy and clear classes. The total training accuracy of the MCM (as compared to MOD35) is 92.94% and the training misclassification rate is 7.06% for the LA PTA from 2002 to 2019. The false positives make up 60.06% of the MCM misclassifications which means the MCM has a higher probability of calling a clear pixel cloudy than a cloudy pixel clear. This is consistent with the lowest performing areas namely snow, bright surfaces, high terrain and sharp surface boundaries such as coastline that are causing false cloudy flags as seen in **Figure 34**.

Table 12: Confusion matrix of the MCM and MOD35 product. Note that the last 6 digits of the numbers have been dropped for readability but still show the relative magnitudes well.

LA PTA 2002-2019	MOD35 Cloudy	MOD35 Clear
MCM Cloudy	28.36% 5.27*10^12 **(True Positive)**	4.24% 7.89*10^11 **(False Positive)**
MCM Clear	2.82% 5.24*10^11 **(False Negative)**	64.58% 1.20*10^13 **(True Negative)**

In the next section individual scenes are inspected by eye to give a practical understanding of the performance and behavior of the MCM in a wide range of cases. This will show the limitations of the mask and how to interpret it for use in the downstream products.

5.2. MCM CLOSER LOOK AND LIMITATIONS

The summary statistics of the MCM show promising performance but the most important aspect of the mask is if it is practically useful when aerosol is present. In this section, individual scenes are inspected showing the MCM, MOD35, a color image, the surface type map and the Distant to Thresholds (DTTs) for each observable. In order to gather the following cases thousands of randomly chosen scenes were inspected by eye to pick representative and extreme examples. The scenes are chosen to highlight the good and the bad parts of the MCM that should be addressed to be able to further improve it and understand its limitations.

Figure 37 is a representative example of the MCM for a cloudy day over the LA PTA, taken Oct. 18th 2015 from Terra MODIS. The color image in the first panel is followed by the MCM, the MOD35 product (Terra MODIS cloud mask) and the surface type map. Following that are the DTT plots of the 7 observables. A description of the color bars is available in the figure caption. On this summer morning there are only a few pixels of snow in the mountains and no sun-

glint water. The first thing to note is the close match between the MCM and the confident-cloudy class for MOD35 both shown in white. This is the target behavior of the thresholds. The DTT for each observable is shown to illustrate the confidence each test has for the cloudy class. Deep red is confident cloudy and deep blue is confident clear. Colors closer to white mean the observation is closer in value to the threshold and therefore less confident. Recall the default cutoff for cloud and not cloud is DTT equal to 0. The 0.65μm BRF DTT shows well defined areas of deep red color over land to denote cloud. It can detect the stratus and the lower cumulus but has a harder time with the thin low clouds in the South East section near LA. Over the desert on the East side the test can distinguish the bright surface from the clouds. Over water the 0.86μm BRF easily detects the bright stratus. Without looking at the rest of the panels these two observables have already detected most of the clouds. The Cirrus DTT does not show much new information. There is not much cirrus present so it's redundantly detecting high, optically thick clouds. The WI is choppy over land suggesting very strict thresholds that allow very little cloud through. However, where it is firing as cloudy, it gets it right. Over water, the WI answer matches closely with the NIR result, however, the cloud edges are less defined as denoted by the red blending into white at the cloud perimeters. The SVI test captures the rich texture of the cloud tops, particularly of the smaller cumulus clouds which are mostly made up of edges at 1km resolution. The NDVI DTT shows how cloud conservative the test is, only calling pixels cloudy over the most optically thick regions. Given many optically thick clouds, the lack of snow, and the lack of sun-glint this would be considered an easy case. A more difficult situation is a mostly clear scene which allows the high texture and reflectance of certain surface types to be falsely flagged as cloudy.

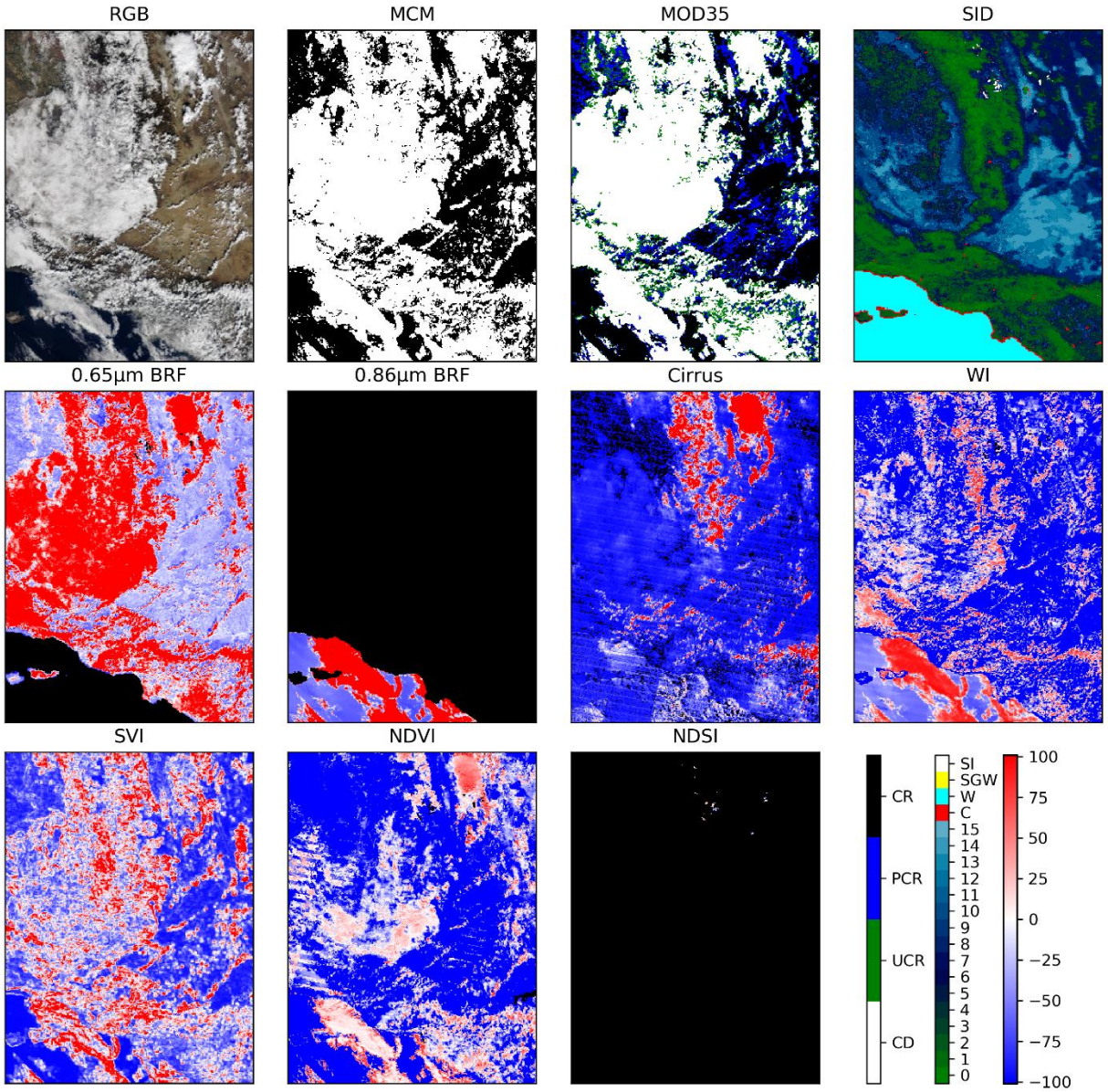

Figure 37: Terra MODIS scene time stamped 2015291.1845 (<year><Julian day>.<UTC time>). The first color bar belongs to the MOD35 panel, where CR is clear, PCR is probably clear, UCR is uncertain clear, and CD is cloudy. The next colorbar is for the SID panel. where SI is snow-ice, SGW is sun-glint water, W is water, C is coast, and 0-15 are the KLID surface types. The last colorbar is the DTT for each observable. In the first three panels, red means no retrival from MODIS. For the last seven panels (DTT), black corresponds to no retrieval from MODIS and/or observable not used over that surface type and/or bad data.

Figure 38 is taken from Terra MODIS over the LA PTA on Dec. 15th, 2018. It is a mostly clear scene with snow in the mountains, no sun-glint and low aerosol loading. Comparing the MCM and MOD35 panels shows how difficult this scene is for cloud masking. Many regions are falsely flagged as cloudy in the MCM including the coastline, the mountain perimeter coincident with the snow edge, and the many bright patches of desert. It should also be noted, however, that MOD35 also has similar issues, and these mistakes will propagate into the MCM's thresholds. The coastline is clear for all the observables except for the NDVI and for the NIR BRF, which both share the 0.86μm band. These errors maybe due to errors in the surface type map. Because the thresholds are derived independently for each surface type, a surface type misclassification, even a small one, will 1) taint the histogram with outlier data and thus skew the threshold calculation and 2) result in situations where the outlier data has the same threshold as the non-outlier data for the same scene type. In this case, parts all along the coastline, including shallow water and land, are effectively labeled as deep ocean. Thus, the thresholds will be tuned for water's weak reflectance in the NIR but the thresholds will not be relevant for the misclassified shallow water and land pixels. A similar story is told by the snow mask. There is an outline of "cloudy" pixels around the snowy Sierra Nevada mountains where snow and ice are labeled as land. This triggers the visible reflectance test which is not designed to be used over snow and ice. Instead, a threshold for a snow-ice free surface types will be applied, causing a false cloudy classification. The NDVI and the WI output lots of false cloudy flags. NDVI tends to mark the most reflective and white desert regions cloudy, mostly found in the North East quadrant of the LA PTA. This suggests the NDVI signature of these surface types is too close to that of clouds, and the histogram threshold method is failing. Thus, for the brightest surface types, the NDVI observable should not be used. The static pattern the WI sprays on the MCM suggests its use over land may not be

practical in clear cases. Finally, the SVI shows false cloud as compared to MOD35 over the mountains, which are highly textured, suggesting that SVI is not a good test for the region. The large areas of blue and green in the MOD35 panel show the uncertainty despite having equipped several thermal channel tests, yet the MCM still offers a close match. The next scene is mostly thin cirrus to illustrate the effects the surface has on thin cloud detection.

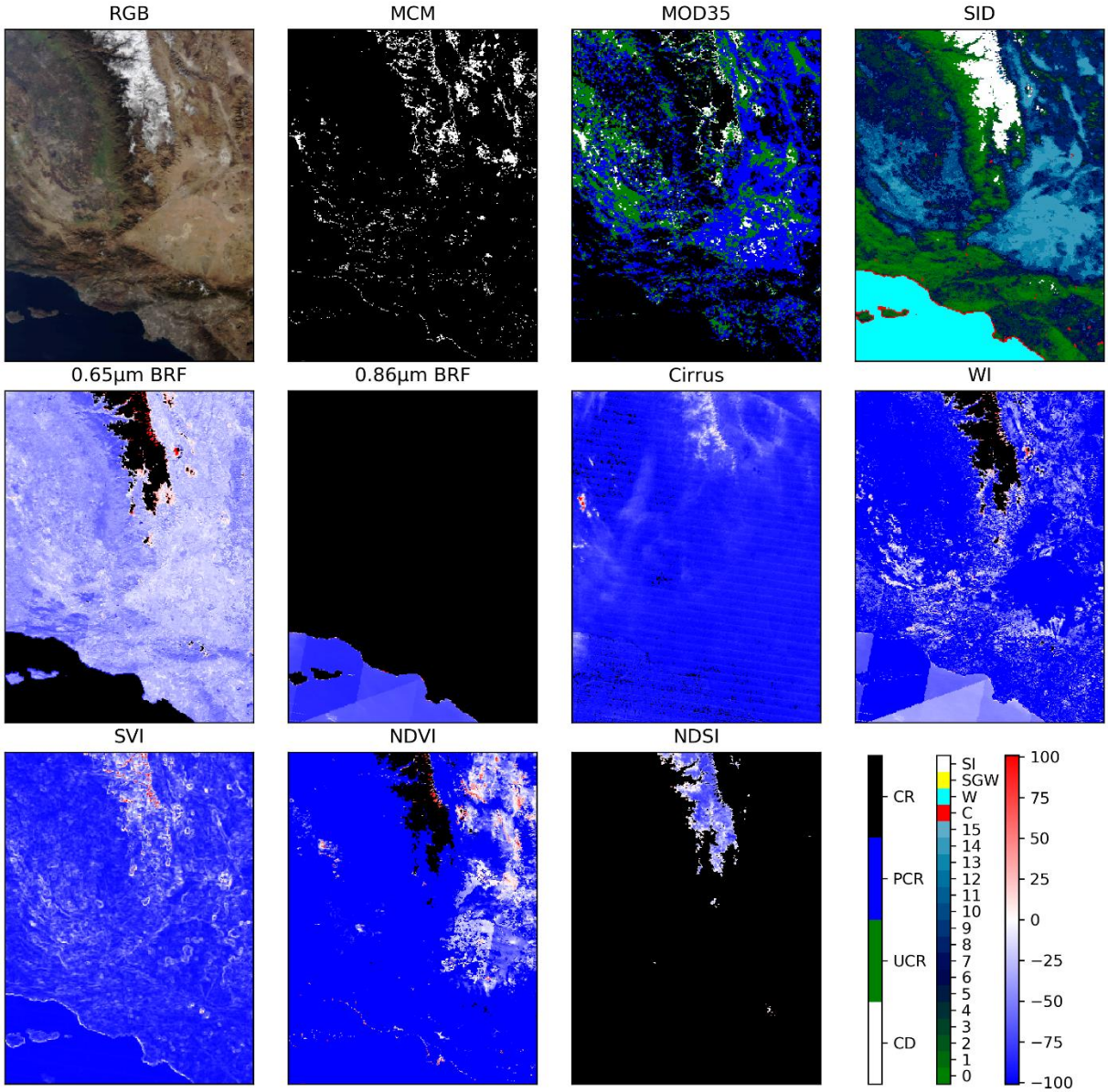

Figure 38: Same as **Figure 37** but for timestamp, 2018349.1830.

Figure 39 shows a predominantly thin cirrus case over the LA PTA on March 25th, 2015, taken from Terra MODIS. Thin cirrus clouds allow a lot of visible radiation reflected off the surface to pass through the cloud and into the sensor. Moreover, the cloud edges are less defined. This makes cloud masking very challenging for this cloud type. The 1.38µm channel can be used to find cirrus clouds more easily (as discussed in **Chapter 2.3.2**) but as seen in the DTT Cirrus panel, a very thin case over desert still allows cloud contaminated pixels to pass as clear. The MCM is able to reproduce most of the cirrus that MOD35 shows with the 1.38µm channel. The missed cloudy pixels can be attributed to the lack of mid IR channels onboard MAIA that MODIS uses to detect thin cirrus in support of the 1.38µm channel (Ackerman et al., 2010). The next case shows the MCM performance over water when strong sun-glint is present.

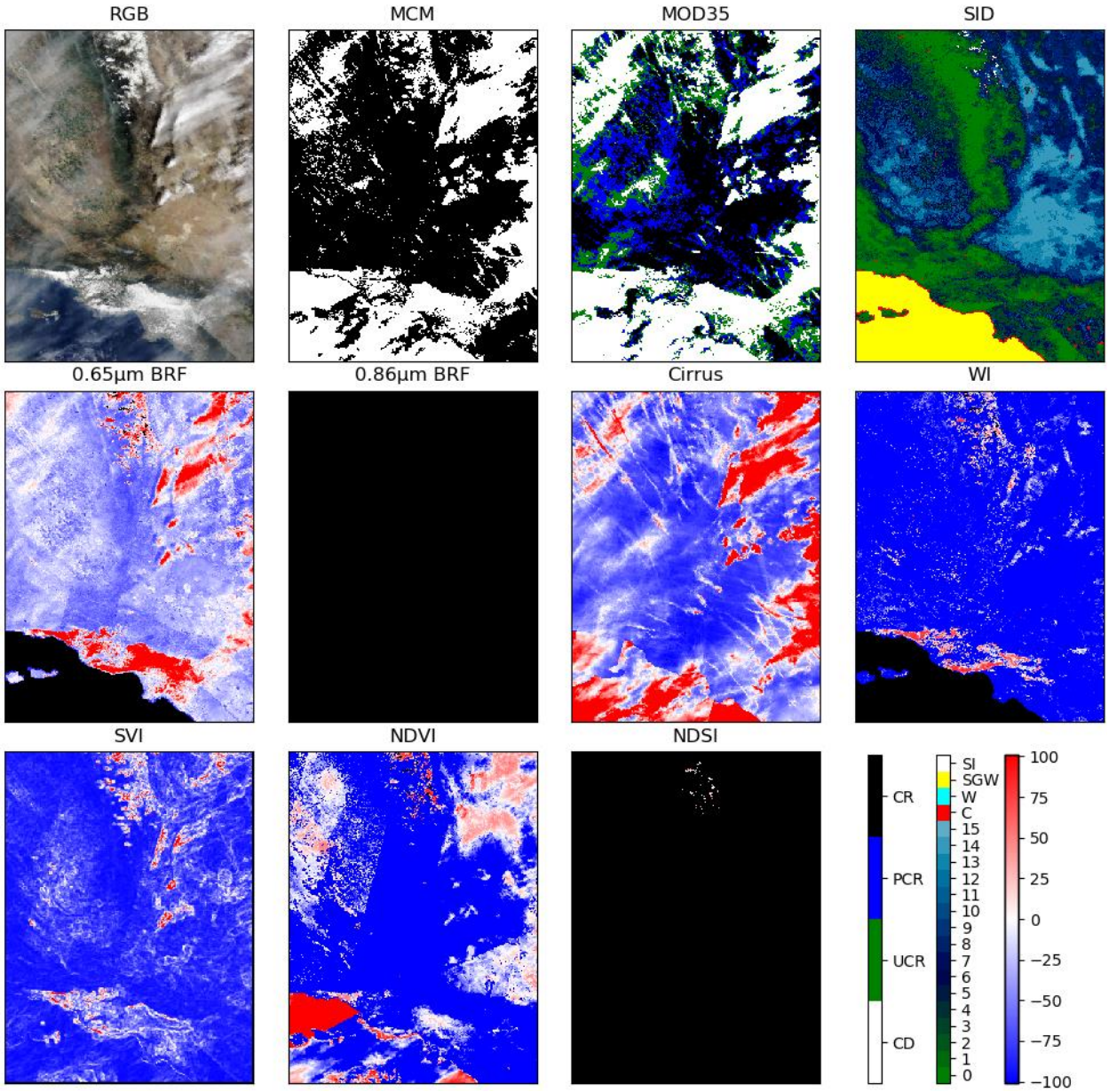

Figure 39: Same as **Figure 37** but for timestamp, 2015084.1820.

Figure 40 is a clear sky case over the LA PTA from Terra MODIS taken on May 13th, 2014. In the South East corner of the LA PTA there is very bright sun-glint. In fact, the whole region is within the 40º scattering angle cone of sun-glint as defined by the MCM. Although MOD35 has trouble classifying this region, it is mostly correct in calling it probably clear (blue shading). The MCM however is calling most of this region cloudy because of the NDVI

observable. In cases of very bright sun-glint it is typical for the NDVI to flag the water as such. Therefore, the NDVI observable should not be used over sun-glint.

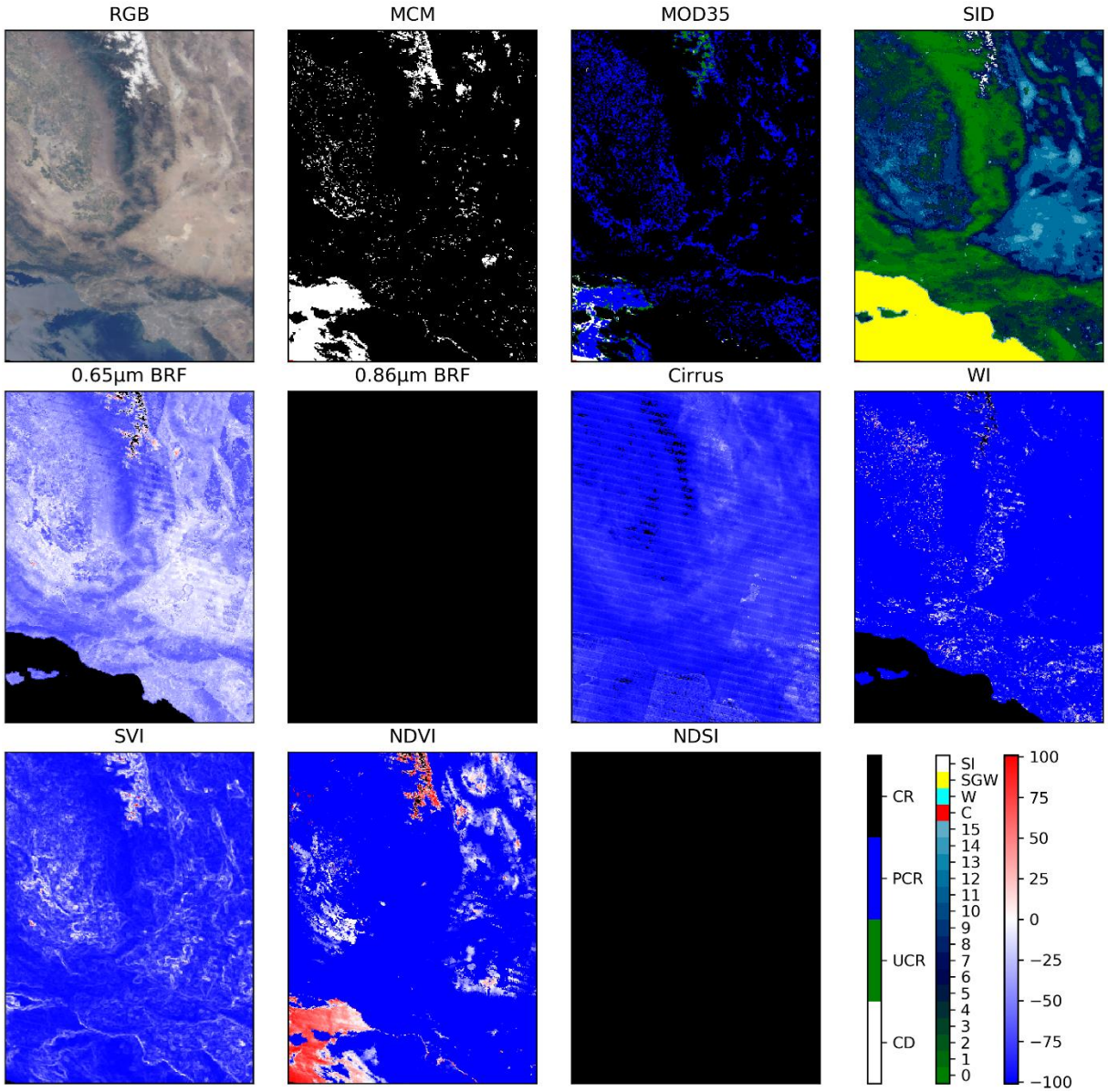

Figure 40: Same as **Figure 37** but for timestamp, 2014133.1900.

The most critical case the MCM can be used for is when there are aerosols present. The cloud conservative thresholds of the MCM will serve to provide as many potentially cloud free pixels to the downstream products as possible. In **Figure 41** there is aerosol present over the LA region and Southern California due to multiple wildfires on the 16th of Dec. 2017. Thick plumes of smoke are visible as well as haze over much of the region. The MCM and MOD35 classifies the thickest plume over water as cloud due to the NIR BRF, the WI, the NDVI and the cirrus observables. This is further evidence that either the WI is not suited for cloud masking, or the thresholds are of poor quality. Because the ocean is very "not white" in the visible wavelengths, the thresholds can be set super high. Meaning that the WI over deep ocean has super high contrast between clouds and the surface. However, that contrast is not strong at all between smoke and clouds. Moreover, it doesn't detect any more cloud than any of the other tests. Similarly, for the NIR BRF, the deep ocean is very dark at this wavelength while clouds remain bright. The thresholds therefore are easily exceeded by the smoke. The NDVI also fails due to the close signature between the smoke and clouds versus the ocean surface. A potential fix for this would be a clear sky restoral operation for smoke, although that is beyond the scope of this thesis. In the North West portion of the PTA, the visible BRF observable is detecting 3 clouds within a layer of smoke. The DTT suggests these clouds, shown in red, have a strong signal against the smoke which is much whiter in color. This is a scenario where the activation value for the visible BRF observable can be increased to mask out the aerosol but preserve the clouds. In fact, the MCM performs quite well outside the main plume of smoke and outperforms MOD35 overall. MOD35 is classifying all the haze as confident-cloud where the MCM is calling it not-cloudy. This illustrates that while the MCM cannot classify the thickest aerosols as clear, it can distinguish between clouds and aerosol haze. The result will be more useful data points for the downstream products vs MOD35.

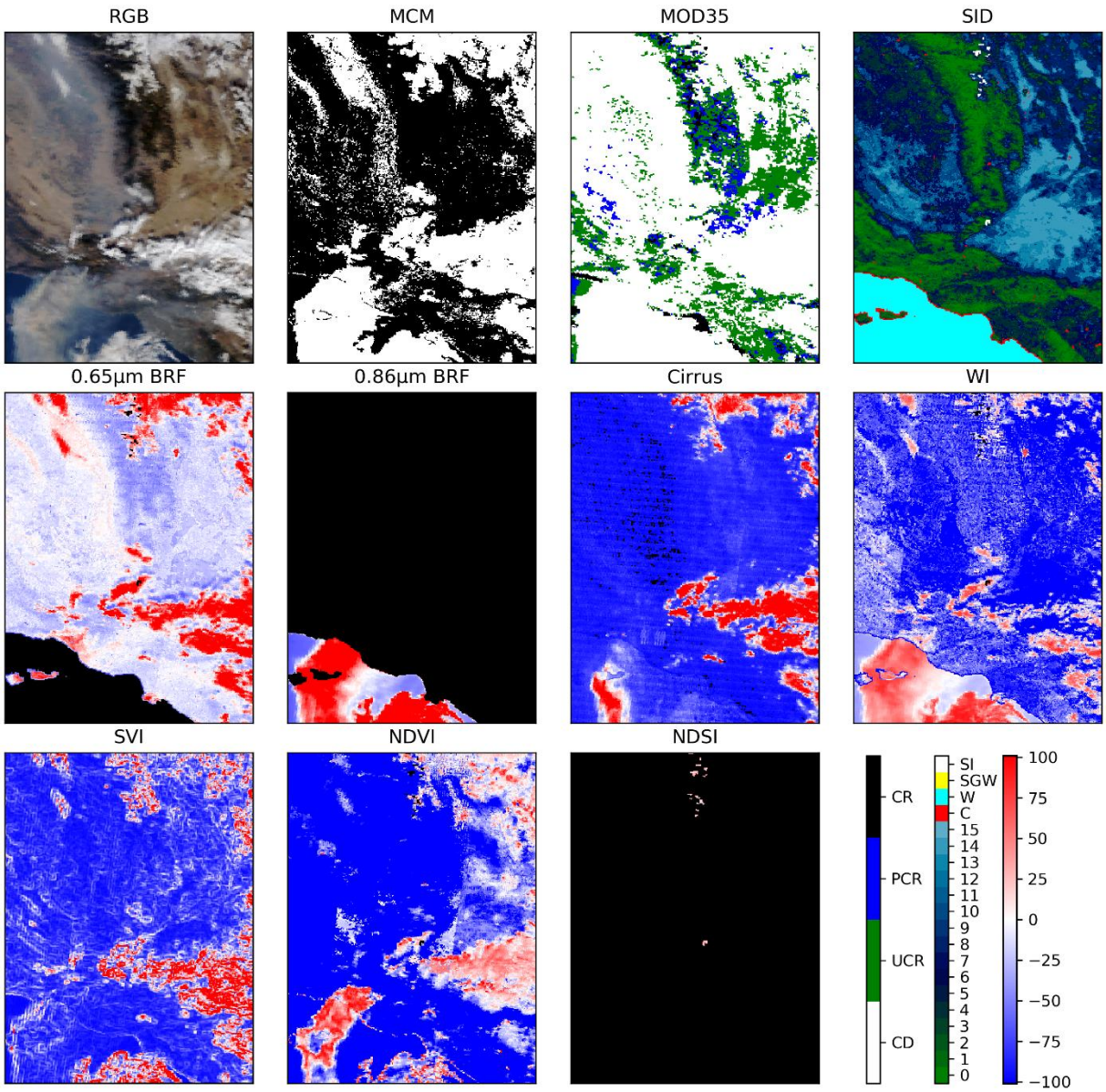

Figure 41: Same as **Figure 37** but for timestamp, 2017350.1805.

In the next and last Chapter, a summary of the algorithm is given followed by the final conclusions about the MCM. Then recommendations are given to improve the mask and to communicate a proper interpretation of the MCM output.

CHAPTER 6 : MCM SUMMARY AND FUTURE WORK

In this thesis the MCM algorithm, its thresholds and its performance are presented in support of MAIA's primary objective, a PM concentration and speciation product. In **Chapter 1** the MAIA mission was presented explaining the need for a day-time cloud conservative cloud mask with tunable thresholds for each of the 11 PTAs. Then the historical framework for cloud detection was laid out to show the MCM's heritage, namely in the Terra MISR RCCM and the Terra MODIS MOD35 cloud masks. In **Chapter 2** the forward pass of the MCM was described with detailed explanations of each function and their purposes, followed by the physical interpretation and utility of each of the 7 observables. They measure brightness and texture, and exploit the absorption properties of water vapor, snow-ice and vegetation relative to clouds. Then the final output of the MCM was shown after the 5 tests are combined using the DTT and AV to define cloudy pixels. Where a pixel is cloudy if the DTT for at least one of the 7 observables is greater than the AV in the configuration file. It was also seen that increasing the AV reduced the cloud fraction at cloud edges and in low confidence areas like coastlines, mountains and snowy terrain. In **Chapter 3** the KLID was defined in support of the thresholds, to classify the surface by maximum BRF. The method used the K-Means clustering algorithm and the max surface BRF from 18 years of MAIAC data to classify the surfaces. Categories 0-15 were assigned to snow-free-land pixels in the PTA grid every 8 days of the Julian calendar at 1 km resolution. The water and coast surface types come from a static land-water mask supplied by the NASA JPL MAIA science team, and the sun-glint mask over water is calculated dynamically for each scene using a scattering angle cone between 0 and 40 degrees. 46 KLID maps, one for every 8-day interval of the year are supplied for each of the 11 PTAs. In **Chapter 4** the pre-launch thresholds were defined for each observable using 18 years of Terra MODIS over passes of the LA PTA. The thresholds were derived to match the

cloud-confident label of MOD35 to by taking the 99th percentile of the combined low middle and high confidence clear values for each observable as the threshold for each scene type. For NDSI and NDVI the mode of the observable for confident cloudy pixels was chosen as the threshold for each scene type. Additionally, the code to produce the thresholds was outlined. Finally, in **Chapter 5** the thresholds were verified and the MCM performance was reviewed. The MCM matched the cloud-confident performance of MODIS over the LA PTA with over 92% accuracy. The MCM is more than 90% accurate over deep non-sun-glint water and land surface types 0-10. The most common mistake was false positives coincident with bright and textured surfaces including the snowy Sierra Nevada mountain tops and the adjacent desert to the east and the coastline. The MCM performs best (>85% accurate) for SZAs between 25 and 53 degrees, VZAs less than 60 degrees and RAAs between 105 and 135 degrees. On a scene-by-scene basis the MCM performed well for optically thick clouds in general, cumulus, stratus and high cirrus and performed poorly for low thin cirrus, thick wildfire smoke, undefined cloud edges and for predominantly clear scenes showing false positives over very bright or textured surfaces.

From the above it is evident that the MCM has several known issues that should be addressed before MAIA launches. 1) The NDVI observable frequently classifies sun-glint water as cloud (as does MOD35). It also falsely classifies bright desert regions and coastlines as cloud. NDVI should therefore not be applied over these surfaces. 2) The cirrus test should not be applied over 2000m above ground level (AGL). Terrain above this level frequently peaks above the moist layer of the atmosphere and is very reflective at 1.38μm BRF and is therefore falsely flagged as cloudy. 3) Moreover, the SVI should not be used over 2000m AGL. Mountains often have SVI on par or greater than clouds causing lots of false cloud flags. 4) The NDSI is only used to detect clouds over snow and snow generally takes on NDSI values of at least 0.4 (Hall et al., 1995), so

therefore all clear scenes will have at least NDSI of 0.4. Clouds will tend to decrease the NDSI therefore the NDSI thresholds should be derived such that the observed value must be less than the threshold to be called cloudy (like the WI) for better performance. 5) Using the Kolmogorov-Smirnov (KS) statistical test, it is shown (**APPENDIX C:**) that the distributions of NDSI thresholds are similar no matter what DOY bin they belong to. This suggests that NDSI thresholds should not be binned by DOY; moreover, the KS technique should be repeated for all observables and for all the dimensions of the threshold dataset to justify the necessity of the threshold binning. This will reduce computation expenses and make the threshold dataset more human interpretable. It will also potentially increase the number of samples in each threshold bin, which could improve the baseline performance of the mask. 6) The WI seems to be superfluous. When clouds are detected by the other observables, WI will also catch them. However, when the scene is mostly clear over land, the WI will flag lots of random pixels as cloud and appear as static across the scene. I recommend the WI not be used over land and potentially not at all, although it requires further investigation to quantify its impact on the MCM. 7) Extremely thick aerosols from wildfires, such as in **Figure 41**, will get flagged as cloudy and it is shown that MOD35 also makes this mistake. Future work needs to reexamine the threshold development with this MOD35 error in mind, since thick aerosols are a critical component of the MAIA mission. Despite these MCM issue, however, it is a great cloud mask as compared to MOD35 and it will serve MAIA's downstream aerosol products well.

REFERENCES

Ackerman, S. A., Frey, R. A., Strabala, K., Liu, Y., Gumley, L. E., Baum, B., & Menzel, P. (2010). Discriminating clear-sky from cloud with MODIS algorithm theoretical basis document (mod35). Madison, Wisconsin: Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin Madison. Retrieved from: https://modis.gsfc.nasa.gov/data/atbd/atbd_mod06.pdf

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C., and Gumley, L. E. (1998). Discriminating clear sky from clouds with MODIS, *J. Geophys. Res.*, *103*(D24), 32141– 32157. https://doi.org/10.1029/1998JD200032

Arthur, D., & Vassilvitskii, S. (2007). *K-means ++: The advantages of careful seeding*. Paper presented at the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, Louisiana.

Bréon, F., & Colzy, S. (1999). Cloud Detection from the Spaceborne POLDER Instrument and Validation against Surface Synoptic Observations. *Journal of Applied Meteorology*, *38*(6), 777-785. https://doi.org/10.1175/1520-0450(1999)038%3C0777:CDFTSP%3E2.0.CO;2

Brimblecombe, P. (1987). *The big smoke: A history of air pollution in London since medieval times.* Methuen, London: Routledge Kegan & Paul.
Goodman, A.H. , Henderson-Sellers, A. (1988). Cloud detection and analysis: A review of recent progress. *Atmospheric Research, 21*(3-4), 203-228. https://doi.org/10.1016/0169-8095(88)90027-0

Chen, T., Rossow, W. B., & Zhang, Y. (2000). Radiative effects of cloud-type variations. *Journal of Climate, 13*(1), 264–286. https://doi.org/10.1175/1520-0442(2000)013%3C0264:REOCTV%3E2.0.CO;2

Choi, H., Bindschadler, R. (2004). Cloud detection in landsat imagery of ice sheets using shadow matching technique and automatic normalized difference snow index threshold value decision. *Remote Sensing of the Environment, 91*(2), 237-242. https://doi.org/10.1016/j.rse.2004.03.007

Chow, J. C. (1995). Measurement methods to determine compliance with ambient air quality standards for suspended particles. *Journal of the Air & Waste Management Association, 45*(5), 320-382. https://doi.org/10.1080/10473289.1995.10467369

Darling, D. A. (1957). The Kolmogorov-Smirnov, Cramer-von Mises Tests. *The Annals of Mathematical Statistics, 28*(4), 823-838. Retrieved from: http://www.jstor.org/stable/2237048

Di Girolamo, L. (2003). Generalizing the definition of the bi-directional reflectance distribution function. *Remote Sensing of the Environment*, 88, 479-482. https://doi.org/10.1016/j.rse.2003.07.004

Di Girolamo, L., Zhan, Y., Zhao, G., Villegas Bravo, J. A., Hong, Y., Lyapustin, A. I., Xu, F., Garay, M. J., Jovanovic, V. M., Diner, D. J. (2019). MAIA level 2 cloud mask algorithm theoretical basis, JPL-103722.

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R. A., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Muller, J. P., Myneni, R. B., Sellers, P. J., Pinty, B., & Verstraete, M. M. (1998). Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview. *IEEE*, *36*(4), *1072-1087.* https://doi.org/10.1109/36.700992

Diner, D. J., Di Girolamo, L., Clothiaux, E. E. (1999). Multi-angle imaging spectroradiometer (misr) level 1 cloud detection algorithm theoretical basis. Retrieved from: ATB_L1CloudDetRevB_Book.fm (nasa.gov)

Diner, D. J., Boland, S. W., Brauer, M., Bruegge, C., Burke, K. A., Chipman, R., Di Girolamo, L., Garay, M. J., Hasheminassab, S., Hyer, E., Jerrett, M., Jovanovic, V., Kalashnikova, O. V., Liu, Y., Lyapustin, A. I., Martin, R. V., Nastan, A., Ostro, B. D., Ritz, B., Schwartz, J., Wang, J., & Xu, F. (2018). Advances in multiangle satellite remote sensing of speciated airborne particulate matter and association with adverse health effects: from MISR to MAIA. *Journal of Applied Remote Sensing, 12*(4). https://doi.org/10.1117/1.JRS.12.042603

Dozier, J. (1989). Spectral signature of alpine snow cover from the landsat thematic mapper. *Remote Sensing of Environment, 28,* 9-22. https://doi.org/10.1016/0034-4257(89)90101-6

Frey, R.A., Ackerman, S.A., Holz, R.E., Dutcher, S., & Griffith, Z. (2020). The Continuity MODIS-VIIRS Cloud Mask. *Remote Sensing*, *12*(20), 3334-3352. https://doi.org/10.3390/rs12203334

Gao, B., Goetz, A. F. H., Wiscombe, W. J. (1993). Cirrus cloud detection from Airborne Imaging Spectrometer data using the 1.38 μm water vapor band. *Geophysical Research Letters, 20*(4), 301-304. https://doi.org/10.1029/93GL00106

Global Burden of Diseases (2017). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the global burden of disease study 2016. *The Lancet*, *390*(10100), 1345-1422. https://doi.org/10.1016/S0140-6736(17)32366-8

Grano, V., Scalione, T., Emch, P.G., Agravante, H., Hauss, B., Jackson, J., Mills, S., Samec, T. K., & Shoucri, M. (2004). End-to-end performance assessment of the National Polar-orbiting Operational Environmental Satellite System environmental data records. *Proceedings of SPIE Weather and Environmental Satellites, 5549.* https://doi.org/10.1117/12.561456

Hall, D. K., Riggs, G. A., Salomonson, V. V. (1995). Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment, 54*(2), 127-140. https://doi.org/10.1016/0034-4257(95)00137-P

Holz, R. E., Ackerman, S. A., Nagle, F. W., Frey, R., Dutcher, S., Kuehn, R. E., Vaughan, M. A., and Baum, B. (2008). Global Moderate Resolution Imaging Spectroradiometer (MODIS) cloud detection and height evaluation using CALIOP. *J. Geophys. Res.*, *113*(D8), D00A19. https://doi.org/10.1029/2008JD009837

Hutchison, K., Iisager, B., & Hauss, B. (2012). The use of global synthetic data for pre-launch tuning of the VIIRS Cloud Mask algorithm. *International Journal of Remote Sensing. 33*(5), 1400-1423. https://doi.org/10.1080/01431161.2011.571299

Jedlovec, G. (2009). Automated detection of clouds in satellite imagery. *Advances in Geoscience and Remote Sensing*, (pp. 303-316). London, United Kingdom: IntechOpen. https://doi.org/10.5772/8326

Jones, A. L., Di Girolamo, L., & Zhao, G. (2012). Reducing the resolution bias in cloud fraction from satellite derived clear-conservative cloud masks. *Journal of Geophysical Research, 117*. https://doi.org/10.1029/2011JD017195

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: trends, perspectives, and prospects. *Science, 349*(6245), 255-260. https://10.1126/science.aaa8415

Kelly, A., Moyer, E., Mantziaras, D., & Case, W. (2014). Terra mission operations: launch to the present (and beyond). Paper presented at the Proceedings of SPIE Earth Observing Systems XIX, 9218, 92180M. https://doi.org/10.1117/12.2061253

Koner, P. K., Harris, A., Maturi, E. (2016). Hybrid cloud and error masking to improve the quality of deterministic satellite sea surface temperature retrieval and data coverage. *Remote Sensing of Environment*, *174*, 266-278. https://doi.org/10.1016/j.rse.2015.12.015

Liang, L., and Di Girolamo, L. (2013). A global analysis on the view-angle dependence of plane-parallel oceanic liquid water cloud optical thickness using data synergy from MISR and MODIS. *J. Geophys. Res. Atmos.*, *118*(5), 2389– 2403, doi:10.1029/2012JD018201.

Loveland, T. R., Belward, A. S. (1997). The international geosphere biosphere programme data and information system global land cover data set (DISCover). *Acta Astronautica, 41*(4-10), 681-689. https://doi.org/10.1016/S0094-5765(98)00050-2

Loeb, N. G., T., Várnai, & Davies, R. (1997). Effect of cloud inhomogeneities on the solar zenith angle dependence of nadir reflectance. *J. Geophys. Res.*, *102*(D8), 9387–9395. https://doi.org/10.1029/96JD03719

Lucht, W., Schaaf, C. B., and Strahler, A. H. (2000). An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE T. Geosci. Remote*, *38*(2), 977–998. https://doi.org/10.1109/36.841980

Luo, Y., Zheng, X., Zhao, T. and Chen, J. (2014). A climatology of aerosol optical depth over China from recent 10 years of MODIS remote sensing data. *International Journal of Climatology, 34*(3), 863-870. https://doi.org/10.1002/joc.3728

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques, 11*(10), 5741–5765. https://doi.org/10.5194/amt-11-5741-2018

Lyapustin, A., Wang, Y., and Frey, R. (2008). An automatic cloud mask algorithm based on time series of MODIS measurements. *J. Geophys. Res.*, *113*(D16), D16207. https://doi.org/10.1029/2007JD009641

Merit, S., & Bruce H. (2009). EVEREST: An end-to-end simulation for assessing the performance of weather data products produced by environmental satellite systems. *Proceedings of SPIE - The International Society for Optical Engineering, 7458*. https://doi.org/10.1117/12.827060

Minnis, P. (1989). Viewing zenith angle dependence of cloudiness determined from coincident GOES East and GOES West data. *J. Geophys. Res.*, *94*(D2), 2303– 2320. https://doi.org/10.1029/JD094iD02p02303

Moody, E. G., King, M. D., & Platnick, S. (2005). Spatially complete global spectral surface albedos: value-added datasets derived from Terra MODIS land products. *IEEE Transactions On Geoscience And Remote Sensing, 43*(1),144–158. https://doi.org/10.1109/TGRS.2004.838359

NOAA-EPA Global Ecosystems Database Project (1992). Global Ecosystems Database Version 1.0 User's Guide, Documentation, Reprints, and Digital Data on CD-ROM. U.S. DOC/NOAA National Geophysical Data Center, Boulder, CO 36 pp.

Pope III, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the Air & Waste Management Association, 56*(6), 709-742. https://doi.org/10.1080/10473289.2006.10464485

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riedi, J. C., Frey, R. A. (2003). *IEEE Transactions on Geoscience and Remote Sensing, 41*(2), 459-473. https://doi.org/10.1109/TGRS.2002.808301

Phulpin, T., Derrien, M. and Brard, A. (1983). A two-dimensional histogram procedure to analyse cloud cover from NOAA satellite high-resolution imagery. *J. Clim. Appl. Meteorol.*, *22*(8), 1332-1345. https://doi.org/10.1175/1520-0450(1983)022%3C1332:ATDHPT%3E2.0.CO;2

Reynolds, D., & Vonder Haar, T. H. (1973). A Comparison of Radar-Determined Cloud Height and Reflected Solar Radiance Measured from the Geosynchronous Satellite ATS-3. *J. Appl. Meteor.*, *12*(6), 1082-1085. https://doi.org/10.1175/15200450(1973)012<1082:ACORDC>2.0.CO;2

Romanov, P. (2017). Global multisensor automated satellite-based snow and ice mapping system (gmasi) for cryosphere monitoring. *Remote Sensing of Environment, 196*, 42-55. https://doi.org/10.1016/j.rse.2017.04.023

Rossow, W. B., & Garder, L. C. (1993). Cloud Detection Using Satellite Measurements of Infrared and Visible Radiances for ISCCP. *Journal of Climate*, *6*(12), 2341-2369. https://doi.org/10.1175/1520-0442(1993)006%3C2341:CDUSMO%3E2.0.CO;2

Saunders, R. W., & Kriebel, K. T. (1988). An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing*, *9*(1), 123-150. https://doi.org/10.1080/01431168808954841

Schiffer, R.A. & Rossow, W.B. (1983). The international satellite cloud climatology project (isccp) the first project of the world climate research program. *Bull. Am. Meteorol. Soc.*, *64*(7),779-784. https://doi.org/10.1175/1520-0477-64.7.779

Shoucri, M., Hauss, B. (2009). Everest: an end-to-end simulation for assessing the performance of weather data products produced by environmental satellite systems. *Proc. SPIE Remote Sensing System Engineering II*, 7458. https://doi.org/10.1117/12.827060

Simmer, C., Raschke, E. and Ruprecht, E. (1982). A method for determination of cloud properties from two-dimensional histograms. *Ann. Meteorol, 20*, 57-69.

Wang, C., Platnick, S., Meyer, K., Zhang, Z., and Zhou, Y. (2020). A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations. *Atmos. Meas. Tech., 13*, 2257–2277. https://doi.org/10.5194/amt-13-2257-2020

Warren, S. G. (1982). Optical properties of snow. *Rev. Geophys.*, 20(1), 67-89. https://doi.org/10.1029/RG020i001p00067

Werner, F., Zhang, Z., Wind, G., Miller, D. J., Platnick, S., & Di Girolamo, L. (2018). Improving cloud optical property retrievals for partly cloudy pixels using coincident higher-resolution single band measurements: A feasibility study using ASTER observations. *Journal of Geophysical Research: Atmospheres*, 123, 12,253-12,276. https://doi.org/10.1029/2018JD028902

Wielicki, B. A., and Parker, L. (1992). On the determination of cloud cover from satellite sensors: The effect of sensor spatial resolution. *J. Geophys. Res.*, 97(D12), 12799– 12823. https://doi.org/10.1029/92JD01061

White, C. H., Heidinger, A. K., and Ackerman, S. A. (2020). Evaluation of viirs neural network cloud detection against current operational cloud masks. *Atmos. Meas. Tech. Discuss. [preprint]*. https://doi.org/10.5194/amt-2020-419

World Health Organization (2013). Health effects of particulate matter: policy implications for countries in eastern europe caucasus and central asia. Retrieved from: http://www.euro.who.int/__data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf

Yang, Y., Di Girolamo, L., Mazzoni, D. (2007). 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*(1–2), 159-171. https://doi.org/10.1016/j.rse.2006.05.020

Yang, Y., & Di Girolamo, L. (2008). Impacts of 3-D radiative effects on satellite cloud detection and their consequences on cloud fraction and aerosol optical depth retrievals. *Journal of Geophysical Research, 113*. https://doi.org/10.1029/2007JD009095

Zhao, G., & Di Girolamo, L. (2004). A Cloud Fraction versus View Angle Technique for Automatic In-Scene Evaluation of the MISR Cloud Mask. *Journal of Applied Meteorology*, *43*(6), 860-869. https://doi.org/10.1175/1520-0450(2004)043%3C0860:ACFVVA%3E2.0.CO;2

Zhao, G., and Di Girolamo, L. (2006). Cloud fraction errors for trade wind cumuli from EOS-Terra instruments, *Geophys. Res. Lett.*, *33*(20), L20802. https://doi.org/10.1029/2006GL027088

Zhang, M. H., et al. (2005), Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements. *J. Geophys. Res.*, 110, D15S02, https://doi.org/10.1029/2004JD005021

Zhu, Z., Woodcock, C.E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment, 118*, 83-94. https://doi.org/10.1016/j.rse.2011.10.028

APPENDIX A: ACRONYMS AND ABBREVIATIONS

AI Artificial Intelligence
AOD Aerosol Optical Depth
AV Activation Value
AVHRR Advanced Very High Resolution Radiometer

BAIR Berkeley AI Research
BRDF Bi-Directional Reflectance Distribution Function
BRF Bi-Directional Reflectance Factor
BT Brightness Temperature

CALIOP Cloud-Aerosol Lidar with Orthogonal Polarization
COD Cloud Optical Depth
CSSC Cloud Screened Surface Classification

DALYs Disability-Adjusted Life Years
DOY Day of Year (Julian calendar)
DTT Distance to Threshold

ECM Enterprise Cloud Mask

FAIR Facebook AI Research

GBD Global Burden of Diseases
GB Giga Byte
GIFOV Ground Instantaneous Field of View
GMASI Global Multisensor Automated Snow and Ice Mapping System

IFOV Instantaneous Field of View
IGBP International Geosphere-Biosphere Programme
INRIA National Institute for Research in Computer Science and Automation
IR Infra-red

JPL Jet Propulsion Laboratory

km kilometers
KS Kolmogorov-Smirnov

LA Los Angeles
LW Long-Wave

MAIA Multi-Angle Imager for Aerosols
MAIAC Multi-Angle Implementation of Atmospheric Correction
MCM MAIA Cloud Mask

APPENDIX A (Cont.)

MISR Multi-Angle Imaging Spectro-Radiometer
µm Micron or micrometer
MODIS Moderate Resolution Imaging Spectro-Radiometer
MVCM MODIS-VIIRS cloud mask

NASA National Aeronautics and Space Administration
NDVI Normalized Difference Vegetation Index
NDSI Normalized Difference Snow Index
NIR Near Infra-red
NIR Ref Near Infra-red Reflectance
NOAA National Oceanic and Atmospheric Administration
NSIDC National Snow and Ice Data Center

OLP Observable Level Parameter

PM Particulate Matter
PTA Primary Target Area

RAA Relative Azimuth Angle
RCCM Radiometric Camera-by-Camera Cloud Mask
RDQI Radiance Data Quality Indicator
RGB Red, Green, Green
RTLS Ross-Thick Li-Sparse

SAA Solar Azimuth Angle
SST Sea Surface Temperature
STA Secondary Target Area
SVI Spatial Variability Index
SW Short-Wave
SZA Solar Zenith Angle

TA Target Area
TB Terra Byte
TOO Targets of Opportunity

VAA Viewing Azimuth Angle
VIIRS Visible-Infrared Imaging Radiometer Suite
Vis Ref Visible Reflectance Test
VZA Viewing Zenith Angle

WI Whiteness Index
WHO World Health Organization

APPENDIX B: CUSTOM VOCABULARY FOR MCM

Observable calculated metric for each pixel used to determine clear or cloudy

Observable Level Parameters/ Bins/ groups/ Scene Type information about the scene type consisting of cos(SZA), VZA, RAA, surface type, and DOY. All quantities discretely binned

Test Determination Step in which the thresholds belonging to a pixel based on its group is retrieved.

Activation Value The confidence a test must give in order to call a pixel cloudy.

APPENDIX C: JUSTIFYING THE BINNING FOR THE THRESHOLD DATASET

The binning of the threshold dataset was mostly based off the MISR RCCM from Diner et al. (1999), but no study was done for the MCM to justify this binning. The binning is used to separate data points by their physical properties in order to optimize thresholds that can distinguish between cloud and not cloud. It is not known if the current resolutions of the bins are optimal for the MCM. This is important to look into because it is 1) computationally expensive to have more bins than needed for processing, 2) if the bins can be combined, the number of samples in the bin will increase, which can lead to an increase in its performance and 3) if more bins are needed to differentiate scene types, that knowledge can be used to expand the threshold dataset. One way to study the first two points is to compare how the distribution of thresholds change from one group of bins to another group of bins. If the change is negligible, then having two different sets of thresholds cannot be justified because they give no new information. For example, if the distribution of thresholds for the NDSI observable are a weak function of the DOY, then why are the NDSI thresholds binned by DOY. This analysis is possible with the use of the Kolmogorov-Smirnov (KS) test which compares 2 empirical distributions of random variables (Darling, 1957). The null hypothesis states that one distribution was sampled from the other and therefore is not unique. If the null hypothesis can be rejected with a p-value lower than 0.05, then the two sets of thresholds must be different and therefore justified. This was done for the NDSI observable, where the distribution of thresholds belonging to DOY bin 0 was compared to DOY bins 1-45 and so on until all unique combinations of DOY bin threshold distributions were compared. The result was that 93.33% of these KS tests could not reject the null hypothesis, meaning the thresholds were in fact a weak function of DOY and therefore the NDSI thresholds should be stratified by much less than 46 DOY bins. This exercise can be

carried out for the sun-view geometry as well. Ideally, that would be done for all 7 observables to cut down the threshold binning to a minimum. This would increase the samples in several bins which can potentially lead to increased performance, and it would cut down on the processing time for the MCM when querying the threshold look up table.

APPENDIX D: MCM CODES

The MCM algorithm code, the threshold derivation code and the threshold analysis code can be found on GitHub here: vllgsbr2/MAIA_CloudMask_Threshold_Development at PTA_agnostic (github.com)

The code for the threshold development is in the “scripts” folder, the code that was delivered to NASA JPL for operational use in the “deliver2JPL” folder and the code use to test and analyze the MCM is in the “test_thresholds” folder. The latest code is in the PTA_Agnostic branch and it is up to date with the data files stored on Keeling. The “test_config.txt” file shows the path to all the data stored on keeling used to create the results shown in this thesis. The operational data sets were reproduced using this code with slight modification form Dr. Guangyu Zhao and are stored in his own directories in Keeling, including the code used to produce KLID.