Studying Anomalous Discrepancies between MODIS and CALIOP Cloud Observations

CyberTraining: Big Data + High-Performance Computing + Atmospheric Sciences

Christine Abraham[1], Olivia Norman[2], Erick Shepherd[2]
Research assistant: Jianyu Zheng[2], Faculty mentor: Zhibo Zhang[2]

[1]Department of Information Systems, University of Maryland, Baltimore County
[2]Department of Physics, University of Maryland, Baltimore County

Technical Report HPCF–2020–12, `hpcf.umbc.edu` > Publications

Abstract

When examining collocated data from the A-Train satellite constellation, there are a notable number of clouds that CALIOP identifies as transparent but MODIS paradoxically reports as having a high cloud optical thickness (COT), therefore implying that the cloud is opaque. We refer to these as "anomalous transparent clouds". Our team is investigating two hypotheses in an effort to explain the occurrence of these anomalies. The first hypothesis is that the anomalies could be MODIS COT retrieval errors due to the misclassification of high albedo surfaces, such as snow and sea ice, as clouds. The other hypothesis is that the anomalies could be clouds which are misidentified as having a high COT due to 3D radiative effects. The former hypothesis was tested by collocating the single-layer cloud anomalies that were over water with NSIDC AMSR-E sea ice observations using a k-nearest neighbors (k-NN) algorithm. We determined that around 50% of such anomalies occur over areas with high sea ice concentrations (95-100%). Further research is required to account for the other half of the anomalies that showed no correlation with high albedo surfaces. We have taken preliminary steps toward exploring whether the cloud 3D radiative effects hypothesis might explain the remaining anomalies.

Key words: MODIS, CALIOP, cloud optical thickness, k-nearest neighbor collocation, 3D radiative effect

1 Introduction

Clouds constantly cover about 60% to 70% of earth's surface. They play an important role in determining the radiative energy balance of Earth and therefore the long-term climate change. It is important to understand the spatial and temporal distributions of clouds as well as their optical and micro-physical properties. At present and in the near future, satellite observation is the only means to observe clouds on a regional to global scale.

The A-Train satellite constellation is a collection of Earth observation satellites which orbit Earth in series and take concurrent measurements of the atmosphere. Two of the

satellites in the A-Train are Aqua and CALIPSO, both of which are equipped with sensors capable of collecting different measurements of cloud properties. Of particular interest to this research is the cloud optical thickness (COT), also known as the cloud optical depth (COD). It is important because COT largely determines how much visible sunlight a cloud would reflect back to space and also how much infrared radiation from the surface a cloud would absorb. Therefore, remote sensing of COT has been the main objective of many satellite sensors.

MODIS, part of Aqua's payload, is an imaging sensor which passively observes the Earth in 36 different spectral bands ranging from infrared to visible violet. The radiation measured in each channel is run through an algorithm to compute information such as the presence and thickness of clouds and aerosols in each pixel.

CALIOP, the CALIPSO payload, uses a lidar signal to measure the atmosphere between the satellite and the ground along a single curvilinear track. The backscattered signal measured by the receiver is processed to determine if there is a cloud present and whether it is transparent or opaque.

As an imaging tool, MODIS collects data in swaths. Since Aqua and CALIPSO orbit synchronously, the CALIOP lidar track is collocated with a portion of the Aqua MODIS swath. These synchronous collocated data afford researchers a unique opportunity to compare the measurements made by both sensors.

Ideally, assuming perfect collocation and synchrony, one should reasonably expect the retrieval results from different sensors to corroborate one another. For instance, if MODIS reports that a particular pixel has a high COT, and is therefore opaque, then CALIOP ought to report that the cloud is opaque. Similarly, CALIOP indicates that a particular cloud is transparent, then MODIS should measure it as optically thin. The same should hold true of other related measurements, such as aerosol optical depth AOD. This lead to the research question: What is the cloud optical thickness retrieved by MODIS for clouds which CALIOP identifies as transparent or opaque?

1.1 Previous Work

Prior to the formation of CyberTraining Team 2, the UMBC *Aerosol, Cloud, Radiation-Observation and Simulation* (ACROS) group addressed this research question by investigating the agreement between MODIS and CALIOP's observations of cloud properties. They compared the COT measurements from MODIS and opacity observations from CALIOP for all Single Layer Liquid-phase clouds (SLC) in 2007. The majority of cloud observations showed agreement between both satellites with around 90% of the transparent clouds having a corresponding COT of less than 10. However, as seen in Figure 1.1 there was an unexpected 2% of clouds that MODIS measured to have high optical thickness (>150), but CALIOP identified as transparent.

Preliminary analysis of these MODIS-CALIOP collocation anomalies showed that there was a positive correlation of the occurrence of these anomalies with mid and high latitudes, and high solar zenith angles (SZA), as seen in Figure 1.2 [4] . Their codebase and collocated

MODIS-CALIPSO-CloudSat dataset served as a foundation for this project.

1.2 Problem Statement

The purpose of this study completed during the CyberTraining program was to investigate possible causes for these "anomalous transparent clouds" to better understand how satellite measurements of clouds could be improved.

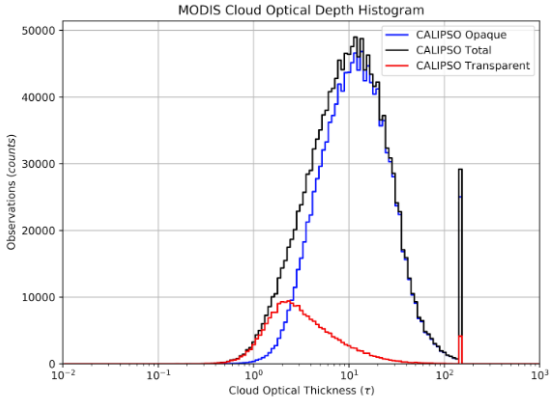


Figure 1.1: Number of observed transparent clouds for different COT values, where all clouds with COT > 150 are grouped together at the tail end of the COT range [4].

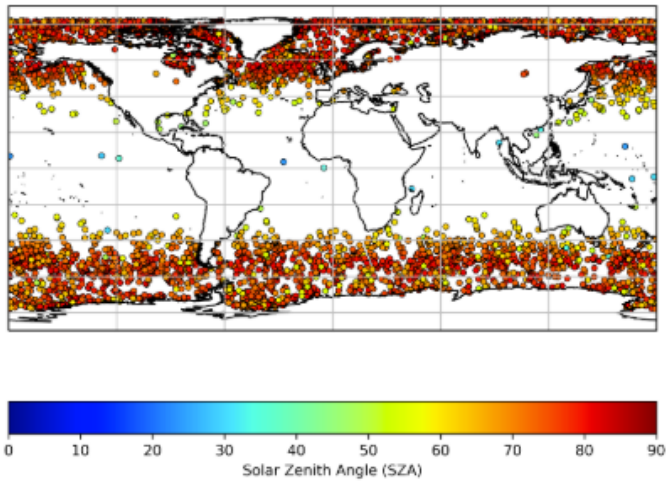


Figure 1.2: Map of the SLC anomalies from 2007 with their color indicating the SZA at the time and location of each anomaly [4].

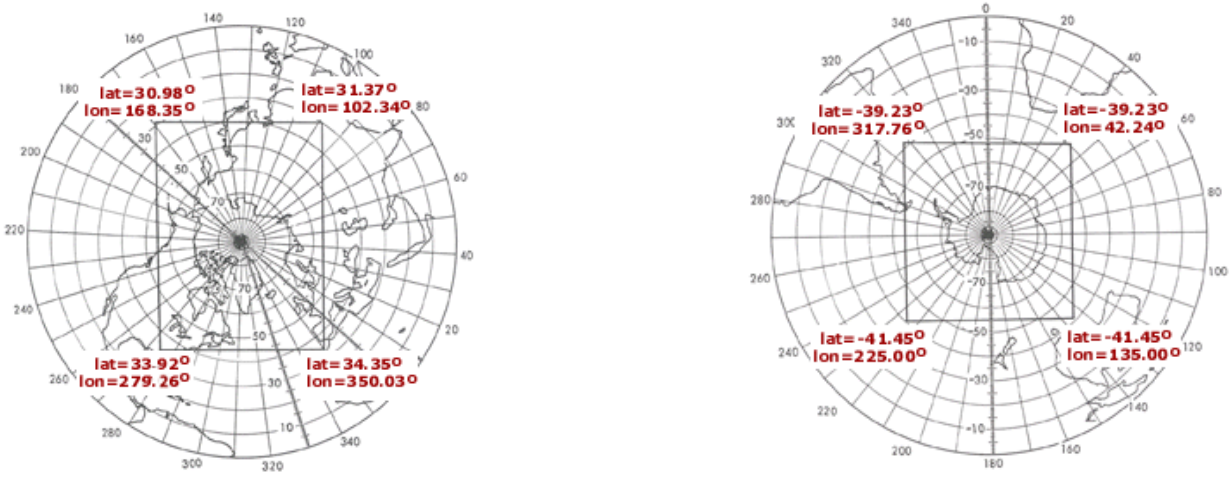
One hypothesis is that the high reflectance of snow and sea ice is misinterpreted as very bright and thick cloud which causes these MODIS COT retrieval errors. This hypothesis is in agreement with the observation that these anomalies mostly occur at high latitudes. The other hypothesis is that the MODIS COT retrieval error is caused by 3D radiative effects. MODIS' algorithms use a 1D theory that does not account for the horizontal properties of clouds. At high solar zenith angles, more light will pass horizontally through the cloud, causing one side of the cloud to be illuminated and the other shadowed. The illuminated side would appear brighter to MODIS and thus the MODIS algorithm may mistakenly treat it as being thicker than it actually is [3]. The method and results found in this study for the sea ice and snow hypothesis and the 3D radiative effects hypothesis are presented in Sections 2 and 3 respectively. The conclusions and plans for future work are presented in Section 4.

2 Snow and Sea Ice

Only the over-water anomaly cases were examined in these study, and therefore the correlation between sea ice and the anomalies was investigated. Possible correlations between the over-land anomalies and snow will be explored in future studies.

2.1 Data

The sea ice data was provided by the AMSR-E sensor on Aqua which provides Level-1A, Level-2A, Level-2B, and Level-3 data products. It measured different variables associated with ocean, atmospheric, and terrestrial processes such as precipitation rate, wind speed, and sea ice concentration, which was the parameter of interest for this study [1]. The daily



(a) Northern Hemisphere

(b) Southern Hemisphere

Figure 2.1: Range of the daily sea ice concentration data collected by the AMSR-E instrument [2]

sea ice concentration (0-100%) is a Level 3 data product, which is calculated using the brightness temperature of the surface in the microwave region as observed by AMSR-E. Daily sea ice concentration data is available for both the northern and southern hemisphere, but is restricted to within a certain range as seen in Figure 2.1. The spatial resolution is 12.5 km and the temporal resolution is one day for the sea ice concentration data. This instrument collected data from June 2002 to October 2011 [2]. The data for 2007 for the northern and southern hemisphere was used in this study.

2.2 Method

The most straightforward way to collocate the NSIDC AMSR-E sea ice concentration data with the coordinates of the MODIS-CALIOP anomalies is to use a naïve brute force algorithm to compute the distance between the position of an arbitrary anomaly and each AMSR-E coordinate. However, this technique is comparatively slow and inefficient, particularly when checking a large number of anomalies against high resolution data. To check the distance of n-many anomaly coordinates from each of m-many AMSR-E datapoint coordinates requires $n \times m$ many comparisons. This naïve brute force approach has an average time complexity of $O(n \times m)$, or $O(n^2)$ with the approximation that $n \approx m$. Moreover, since the coordinates are geodetic, computing the distance between any two points requires the use of multiple trigonometric functions, which are computationally expensive compared to the arithmetic distance formula used in Cartesian space: For an $O(n^2)$ complexity algorithm, the difference is especially noticeable.

With the limited efficiency of the naïve brute force approach in mind, we sought to develop a more efficient collocation algorithm. Our first attempt involved averaging the AMSR-E sea ice concentrations into $0.25^\circ \times 0.25^\circ$ resolution cells to form a grid. In this form, the problem of collocation reduces to identifying which cell contains the anomaly. While computationally efficient, this approach presents several obvious issues, particularly when applied to geodetic point coordinates. Precision is inherently lost during the gridding and averaging in any coordinate system, but in spherical coordinates, the precision of the collocation is further reduced the farther the anomaly is from either pole due to the increasing area of the grid cells. Consequently, the gridded averages collocation approach proved to be a poor candidate for our analysis.

In an effort to improve upon the efficiency limitations of the naïve brute force approach and the precision loss of the gridded averages approach, we instead used the k-nearest neighbors (k-NN) algorithm, an instance-based machine learning algorithm which builds a k-dimensional (k-d) tree to efficiently locate the AMSR-E point coordinate closest to an arbitrary MODIS-CALIOP anomaly. Because it minimizes the distance between point coordinates, it does not suffer the precision loss exhibited by the gridded averaging. Furthermore, k-d tree reduces the number of comparisons needed to identify the nearest point versus the brute force approach. Where the naïve brute force algorithm has an average time complexity of $O(n \times m)$, training the k-NN model has $O(m \log m)$ complexity and traversing the k-d tree for n anomalies has $O(n \log m)$ complexity. By summing and again making the approximation that $n \approx m$, this is an improvement from $O(n^2)$ to $O(2n \log n)$ complexity.

The NSIDC AMSR-E product is assembled using an EASE-Grid where the given geodetic point coordinates are the centers of each cell. Since the k-d tree for the k-NN algorithm against which the anomaly coordinates are checked is constructed using these gridpoints, which are common between all data files in the product, in practice, the model only needs to be trained once. Once generated, it can be saved and reused in subsequent collocations, reducing the time taken to train the model to that needed to read it from disk. The result is an improvement in the run-time complexity of the k-NN algorithm over the naïve brute force algorithm from $O(n^2)$ to $O(n \log n)$ complexity.

We were able to further improve the efficiency of the collocation through judicious use of geographical distance formulae. When computing geographical distances, there are a number of formulae available. Two formulae of note are the Haversine formula and Lambert's formula for long lines. Lambert's formula is the most accurate, on the order of $10m$. However, the Haversine formula is less computationally expensive. When applying our k-NN model, we used the Haversine formula to identify our k nearest neighbors, then applied Lambert's formula to compute the distances of the anomaly from this reduced set of gridpoints and identify which is the closest. This allowed us to retain both the speed of the Haversine formula and the high accuracy of Lambert's formula. If the final Lambert distance of anomaly from its nearest AMSR-E gridpoint was greater than would produce an overlap of the center of the CALIPSO lidar shot with the edge of an AMSR-E EASE-Grid cell, the anomaly was considered uncollocated and its associated sea ice concentration was assigned a null value. The scale of the anomaly relative to the collocated EASE-Grid cell, as well as an example of a properly collocated case, are illustrated in Figure 2.2.

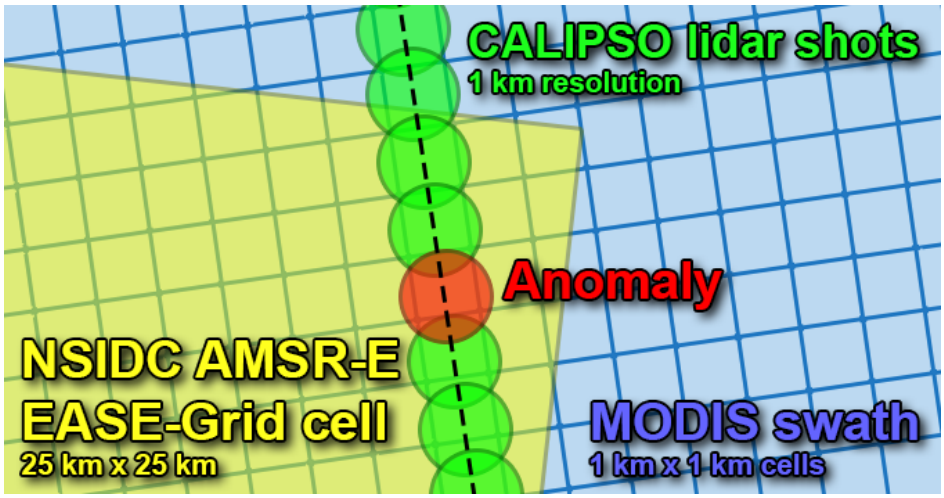


Figure 2.2: A depiction of the overlapping anomaly and NSIDC data points used in the collocation of these two data sets using the k-NN algorithm

2.3 Results

The results for the collocation of the 7236 SLC over-water anomaly cases in 2007 with the sea ice data are presented in Figure 2.3. Around 52% of the anomalies occurred over

high sea ice concentrations (95-100%), while around 26% of these anomalies occurred over very minimal to no sea ice (0-5% concentration). Around 18% of the anomalies were unable to be collocated with the sea ice data. These uncollocated anomalies are those that either did not fall within the range of the AMSR-E data product, as seen in Figure 2.1, or which were not close enough to any AMSR-E gridpoint to make an accurate collocation. Figure 2.4 shows the location and sea ice concentration of all of the anomalies that were able to be collocated successfully.

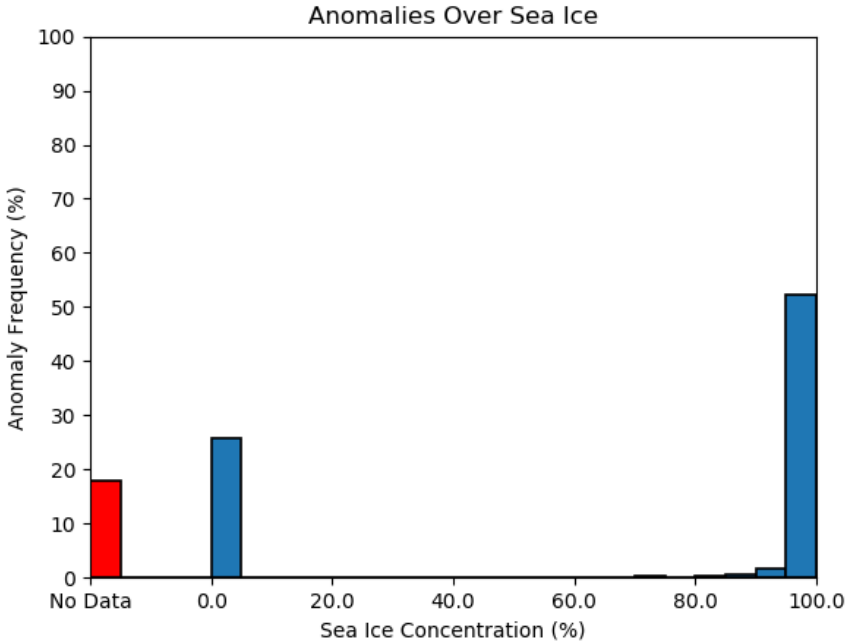


Figure 2.3: Percent of anomalies with various sea ice concentrations (0-100% in intervals of 5%) and the percentage of the anomalies that were unable to be collocated (reported as "No Data")

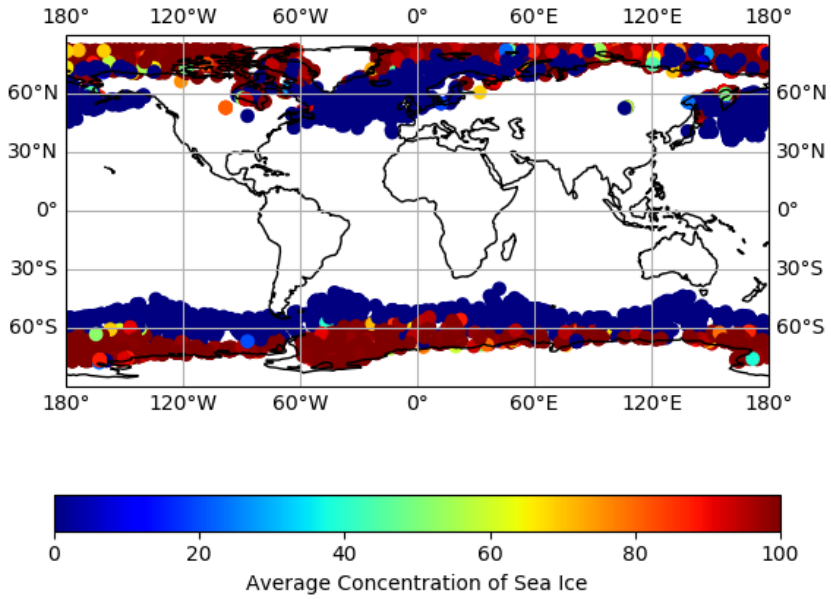


Figure 2.4: Map of the anomalies with their respective collocated sea ice concentration values

3 3D Radiative Effects

3.1 Method

If the anomalies occur on the illuminated side of the cloud, this would provide support for the 3D radiative effects hypothesis. To determine where on the cloud the anomalies are located, the slope of the cloud and direction of sun at the point of the anomaly needs to be determined.

During the Cybertraining program, the slope was calculated using the MODIS' cloud top heights. The CALIOP's layer top heights were also accessed, but the slope was not officially calculated yet. For the case of CALIOP's layer top heights, the relevant layers needed to be identified as clouds in order to be included.

For all of the SLC anomalies in 2007, the cloud top height for the anomaly and for the +/- 5 MODIS pixels surrounding the anomaly were found. Using all 11 points, the slope was then calculated using linear regression with Numpy's polyfit with $deg = 1$. This resulting slope was not a good approximation for the slope at the point of the anomaly and so another method has to be developed to determine the slope.

3.2 Results

The MODIS and CALIOP cloud top height along the track was plotted. An example of such profile plot is give in Figure 3.1. This specific anomaly is one that is not acounted for by the sea ice hypothesis since it occurs over 0% sea ice concentration. For this anomaly, the SZA and solar azimuth angle (SAA) were read in from the CALIOP data file and are 63.77° and 283.64° respectively. These angles define the direction of the sun. Using the direction of the satellite, which is provided in map form along with images of the CALIPSO lidar profile plots available online, and the direction of the sun, it can be determined that the sunbeam is coming from the opposite direction that the satellite is moving. In other words, the sunbeam is coming from the left side of the plot below. This means that the anomaly would be on the illuminated side of a cloud, which would provide support for the 3D radiative effects hypothesis.

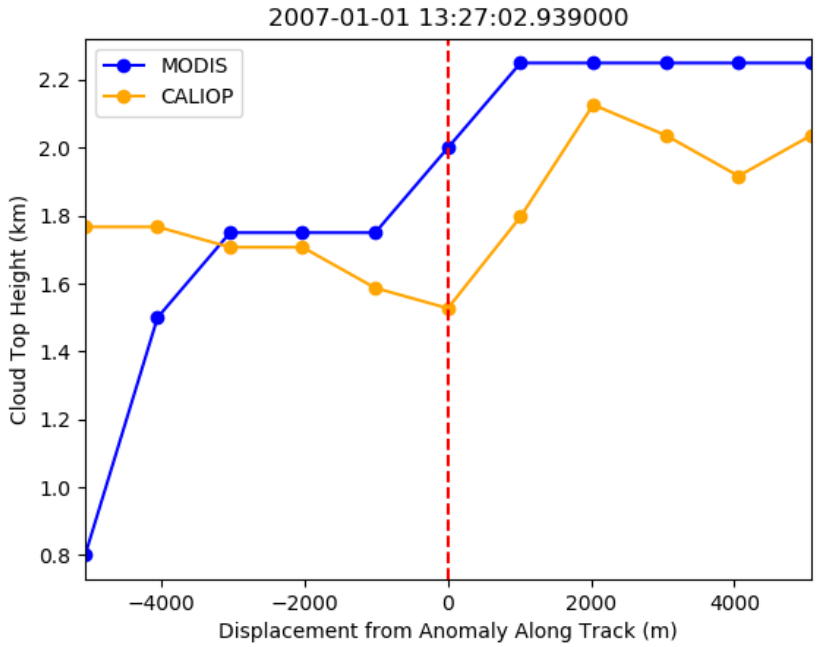


Figure 3.1: MODIS and CALIOP cloud top height for points surrounding the anomaly, which is marked with the dashed red line. The SZA and SAA measured by CALIOP at the point of the anomaly indicate the sunlight is coming from what equates to the left side of this figure, which means that the anomaly is illuminated.

4 Conclusions and Future Work

There appears to be a strong correlation between sea ice and the anomalies, with over 50% of the anomalies occurring over high sea ice concentrations. This hypothesis will be further explored by expanding the data set beyond 2007. Also, the SLC over-land anomalies, which

occur more frequently than the over-water cases, will be tested to see if there is a correlation with snow. Once a sufficient data set for snow has been located, the same collocation processes used for the sea ice and anomaly data can be utilized.

However, there are still about 50% of the SLC over-water anomalies cases in 2007 that are unaccounted for by the sea ice hypothesis. These anomalies could be related to 3D radiative effects, which will be the focus of subsequent research studies. The process of determining if the anomaly is on the illuminated side of the cloud will be automated in order to see if there are a number of cases, especially ones not over ice, that occur on the illuminated side of the cloud. This can be determined by finding the dot product of the slope vector and the sunbeam vector, the sign of which indicates whether the anomaly is on the illuminated side or not.

Acknowledgments

This work is supported by the grant "CyberTraining: DSE: Cross-Training of Researchers in Computing, Applied Mathematics and Atmospheric Sciences using Advanced Cyberinfrastructure Resources" from the National Science Foundation (grant no. OAC–1730250).

The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS–0821258, CNS–1228778, and OAC–1726023) and the SCREMS program (grant no. DMS–0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC). See `hpcf.umbc.edu` for more information on HPCF and the projects using its resources.

Glossary

A-Train The Afternoon Train satellite constellation.

lidar Light detection and ranging.

Acronyms

AMSR-E Advanced Microwave Scanning Radiometer for EOS.

AOD Aerosol Optical Depth.

CALIOP Cloud-Aerosol Lidar with Orthogonal Polarization.

CALIPSO Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations.

COT Cloud Optical Thickness.

MODIS Moderate Resolution Imaging Spectroradiometer.

References

[1] Donald J. Cavalieri, Thorsten M. Markus, and Josefino C. Comiso. *AMSR-E Overview*. 2020. URL: https://nsidc.org/data/amsre.

[2] Donald J. Cavalieri, Thorsten M. Markus, and Josefino C. Comiso. *AMSR-E/Aqua Daily L3 12.5 km Brightness Temperature, Sea Ice Concentration, & Snow Depth Polar Grids, Version 3*. Boulder, Colorado, 2014. DOI: https://doi.org/10.5067/AMSR-E/AE{_}SI12.00. URL: https://nsidc.org/data/AE_SI12/versions/3.

[3] Tamás Várnai and Alexander Marshak. "Observations of three-dimensional radiative effects that influence MODIS cloud optical thickness retrievals". In: *Journal of the Atmospheric Sciences* 59.9 (May 2002), pp. 1607–1618. ISSN: 00224928. DOI: 10.1175/1520-0469(2002)059<1607:OOTDRE>2.0.CO;2. URL: https://doi.org/10.1175/1520-0469(2002)059%3C1607:OOTDRE%3E2.0.CO%20http://0.0.0.2.

[4] Zhibo Zhang et al. *Understanding the Quantitative Connection Between Cloud Opacity and Cloud Optical Thickness Using CALIOP and MODIS Observations*. Poster session presented at: MODIS/VIIRS Science Team Meeting. 6th Annual Conference of the MODIS and VIIRS Science Teams. College Park, MD, Nov. 2019.