Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Studying Anomalous Discrepancies between MODIS and CALIOP Cloud Observations

CyberTraining: Big Data + High-Performance Computing + Atmospheric Sciences, Spring 2020

Christine Abraham[1], Olivia Norman[2], Erick Shepherd[2]

Faculty advisor: Dr. Zhibo Zhang[2]
Research assistant: Jianyu Zheng[2]

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

Acknowledgements: NSF, UMBC, HPCF

Project Background

Earth Observation Satellites

Terra & **Aqua**

- **MODIS** (The Moderate Resolution Imaging Spectroradiometer)

CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations)

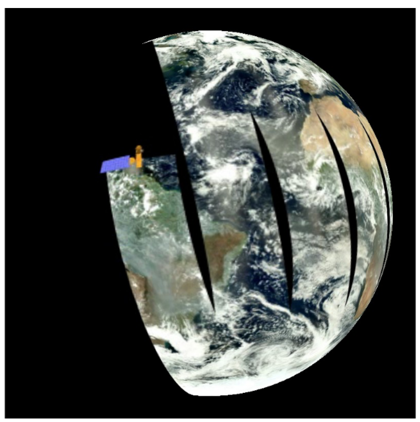
- **CALIOP** (Cloud-Aerosol Lidar with Orthogonal Polarization)

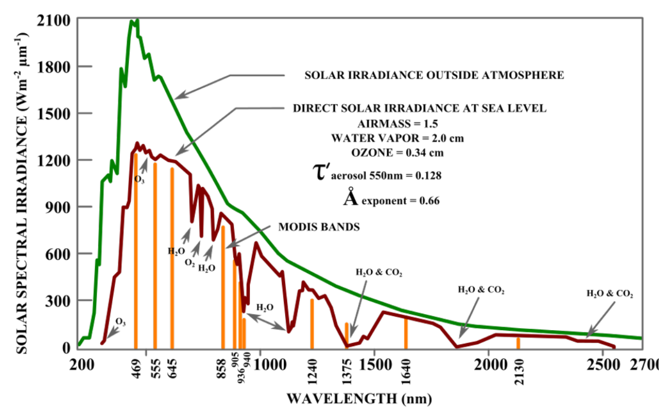
A-train (Afternoon Train)

- Satellite constellation including Aqua with a MODIS payload and CALIPSO with the CALIOP payload
- Flying in series and taking concurrent collocated measurements of Earth's atmosphere

MODIS

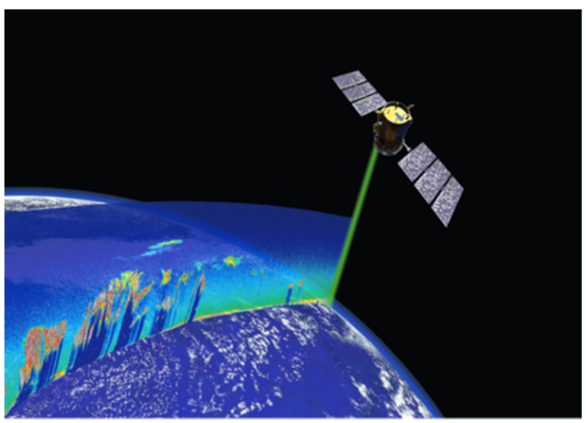
- Uses an imaging sensor to image the Earth in swaths in 36 different bandwidths of light ranging from infrared to visible violet.
- Metrics such as the presence and thickness of clouds are inferred from an algorithm using data such as brightness and spatial variability.

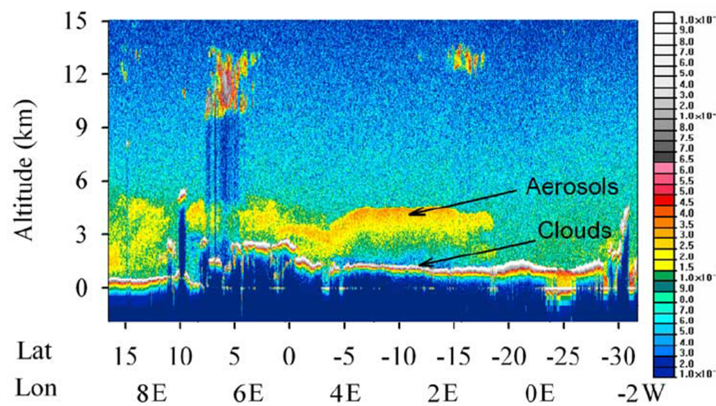




CALIOP

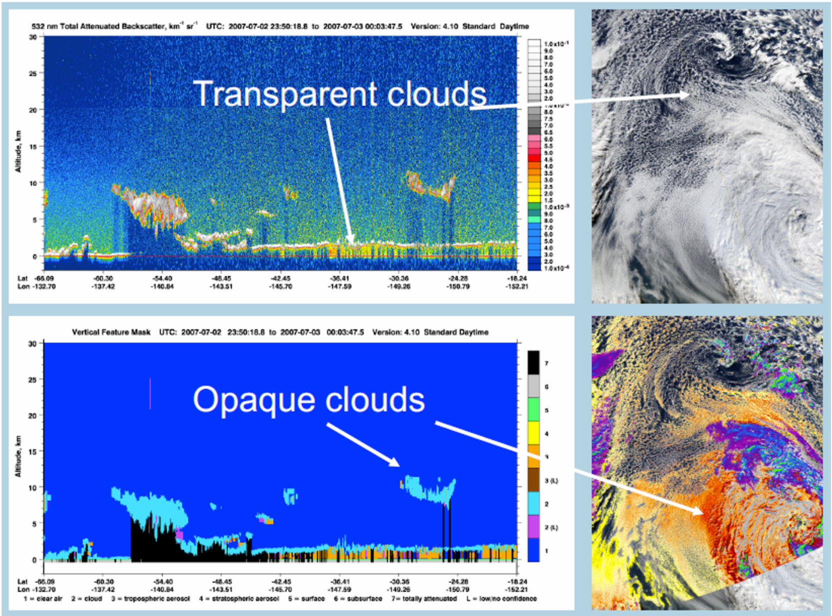
- Uses a lidar (light detection and ranging) signal to measure a single track of the atmosphere between the satellite and the ground.
- Infers the presence and properties of clouds and aerosols from the depolarization ratio and backscatter signal received by the satellite and is calibrated against surface sun photometers (e.g., AERONET).





Collocated MODIS / CALIOP Observations

- The MODIS pixels supply the cloud optical thickness (COT, also called “cloud optical depth” or “COD”) and aerosol optical thickness (AOT or AOD).
- The CALIOP track supplies layer data and information about aerosol size and composition, cloud phase, and whether a cloud is transparent or opaque.



Prior Work

Identifying Anomalies

Through preliminary work, our team has already answered the first half of our research question:

Do MODIS' COT measurements and CALIOP's opacity observations agree?

For the most part, yes: the measurements made by both instruments behave as one would hope more often than not. **However, we have identified a statistically significant set of anomalies which warrant further investigation.**

Identifying Anomalies

These anomalies fall into two broad classes:

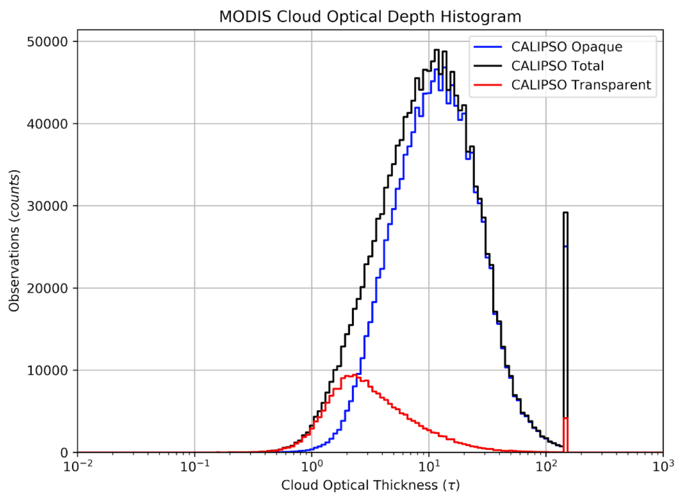
- **A cloud is *transparent* per CALIOP but has a *high COT* per MODIS.**
- **A cloud is *opaque* per CALIOP but has a *low COT* per MODIS.**

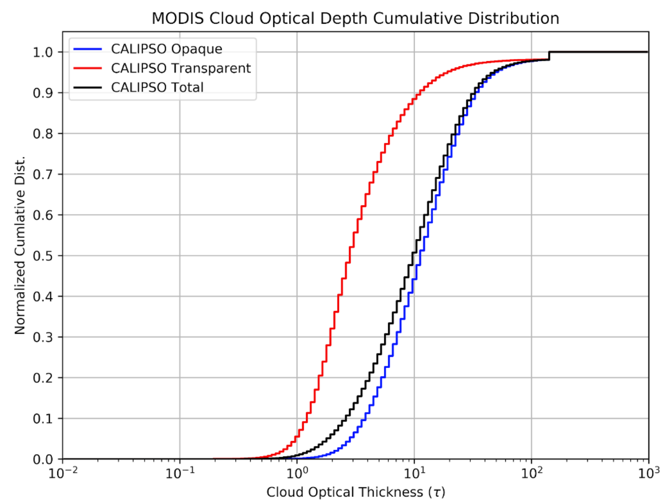
In order to isolate the simplest cases of such anomalies, we restricted our collocated data exclusively to instances of **single-layer water phase clouds over water** and resolved to identify the first case of **transparent** clouds per CALIOP with **high COT** per MODIS.

COT of Transparent Clouds

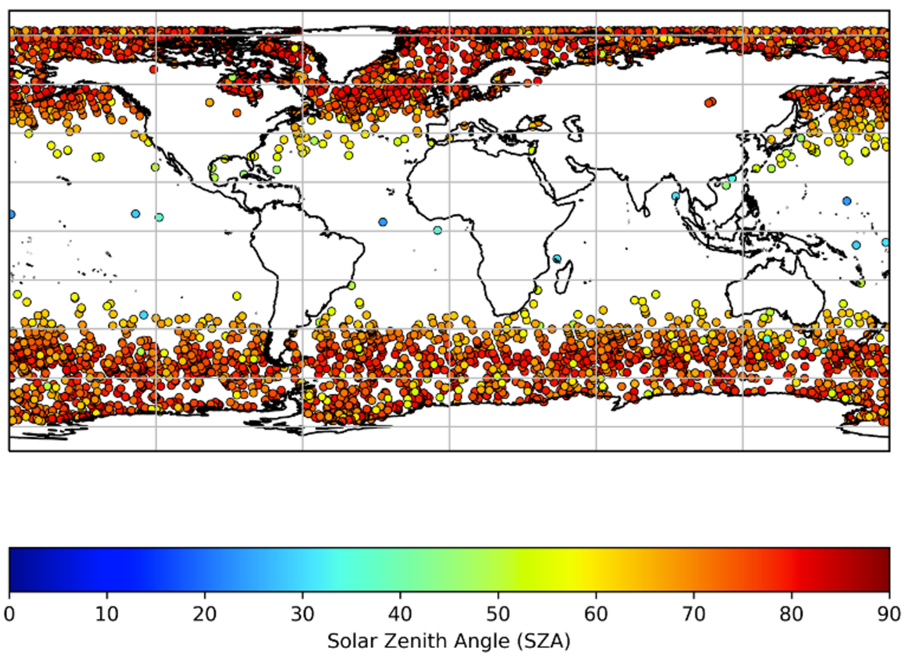
Mean: 8.0
Mode: 2.3
Median: 3.0
τ < 10 ≈ 90%
τ > 100 ≈ 2%

5% - 95% range:
$1.0 < \tau < 20.7$





However, there are a notable number of identified **thick transparent clouds**.



A first step in identifying whether there is a pattern to these **thick transparent cloud** anomalies is identifying correlations.

There appears to be a strong latitudinal correlation. In terms of cloud radiative effects, this in turn correlates to a high solar zenith angle (SZA)..

Problem Definition

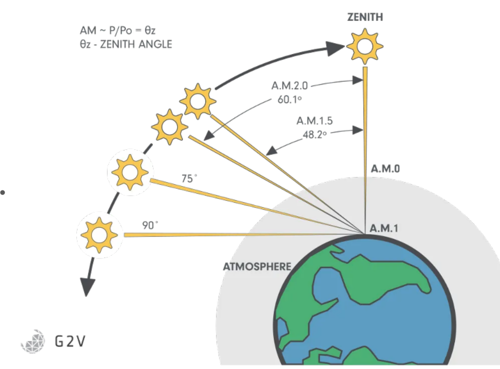
Examining Possible Explanations

What are the potential causes of the "anomalous transparent clouds"?

- **Hypothesis #1:** MODIS/CALIOP collocation error.
  - Unlikely, as this does not explain the dependence on solar zenith angle.
- **Hypothesis #2:** CALIOP opacity retrieval error.
  - Unlikely, as this does not explain the dependence on solar zenith angle.
- **Hypothesis #3:** MODIS COT retrieval error over snow and/or sea ice.
  - Possible, as both clouds and sea ice have high reflectance and low spatial variability.
- **Hypothesis #4:** MODIS COT retrieval error caused by 3D radiative effects.
  - Possible, as clouds illuminated by 3D radiative effects may appear brighter, and therefore thicker, to the MODIS classification algorithm.

Expounding Hypothesis #4: 3D Radiative Effects

The algorithms to infer **COT from spectral imaging data use a 1D theory** which **ignores the horizontal properties of clouds**. It is conceivable that at extreme solar zenith angles, when a cloud is illuminated from the side, more light reflects out of the top of the cloud than when illuminated from lower solar zenith angles. This results in the **cloud appearing brighter** than it otherwise would, and thus **the MODIS algorithm may mistakenly treat it as being thicker than it actually is.**



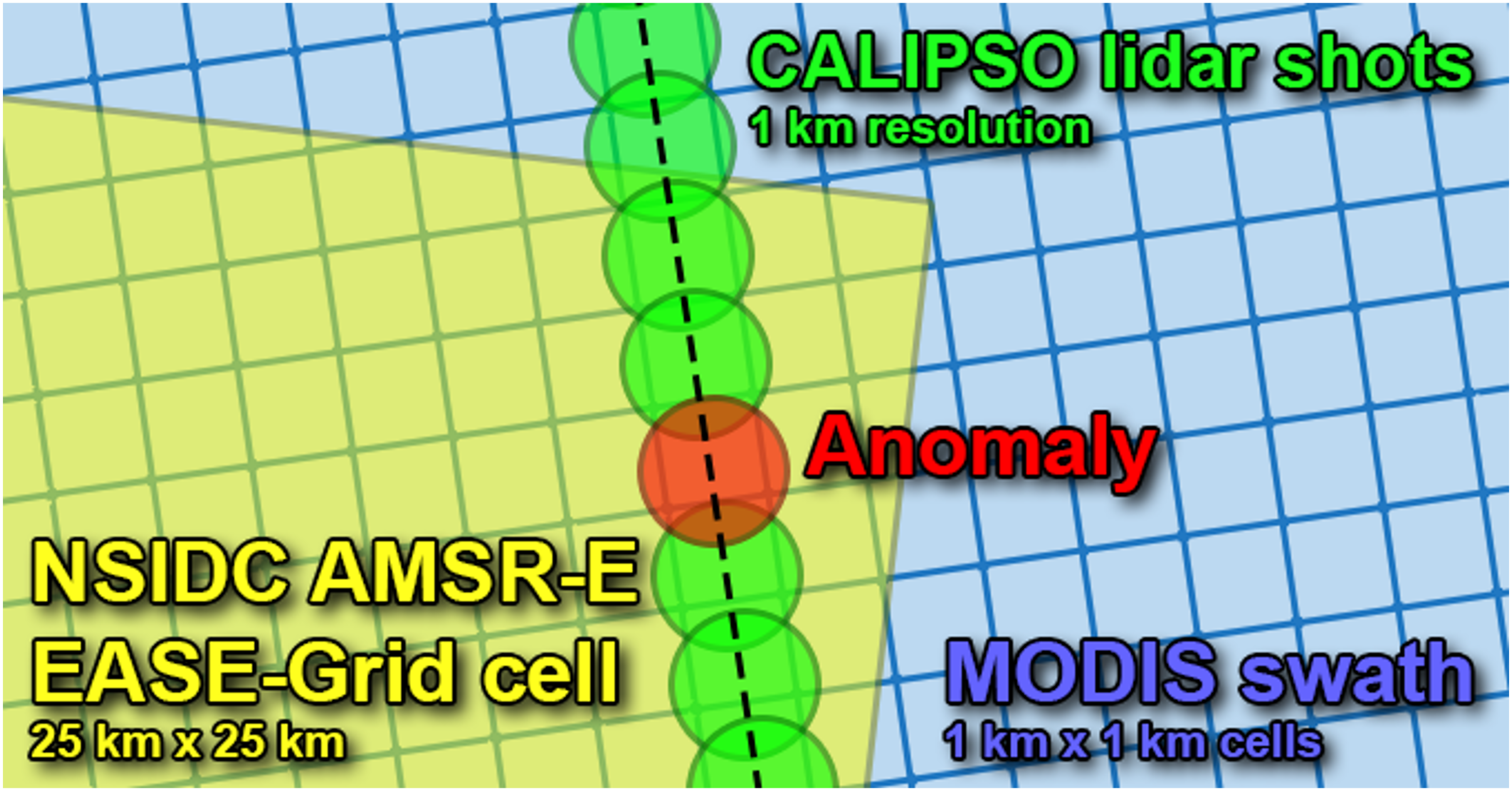
Research Questions

- Does the sea ice hypothesis explain any of the anomalies?
- Does the 3D radiative effects hypothesis explain any of the anomalies?

Sea Ice Hypothesis

Approach

- We want to determine if there is a correlation between the location of of *single-layer water-phase cloud anomalies over water in 2007* and the concentration of sea ice.
- We are using a sea ice data set from NSIDC (https://nsidc.org/data/ae_si12)
- We collocated sea ice and anomaly data using the nearest overlapping neighbor


CALIPSO lidar shots
1 km resolution
Anomaly
NSIDC AMSR-E
EASE-Grid cell
25 km x 25 km
MODIS swath
1 km x 1 km cells

Nearest Neighbor - Brute Force

In order to identify the NSIDC AMSR observation collocated with a given anomaly, we needed to find the nearest neighbor. One way to do this is a naive **brute force algorithm**, where we look at every NSIDC AMSR coordinate, check its distance from the anomaly, and find the minimum distance. However, for the surface of a 2D spheroid such as the Earth, this has $O(n^2)$ complexity, which makes it comparatively slow.

Nearest Neighbor - k-Nearest Neighbor

A more efficient way to collocate the points is the **k-nearest neighbor algorithm**, which takes all of the NSIDC AMSR coordinates and constructs a **k-d tree**. Building this model has ***O(n ln(n))*** complexity, and searching for the nearest neighbor of a point in the tree has ***O(ln(n))*** complexity. This makes it more efficient than the naive **brute force algorithm**.

Nearest Neighbor - k-Nearest Neighbor

We achieved an even greater boost to efficiency by recognizing that the observation coordinates for the NSIDC AMSR dataset is identical between files. This means that only two **k-d trees** need to be constructed: one for the northern hemisphere and one for the southern hemisphere. By only generating each of these models once and saving the resulting Python **k-d tree** object as a **pickle** file, the ***O(n ln(n))*** complexity required to generate the **k-d tree** can be circumvented entirely for all subsequent collocation runs. Thus, we reduced the collocation run-time from ***O(n^2)*** to ***O(ln(n))*** complexity.

Shortest Distance

To identify the distance between two ***(latitude, longitude)*** pairs on a sphere, we used the **Haversine formula**:

$$d = 2r \arcsin\left(\sqrt{\operatorname{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\operatorname{hav}(\lambda_2 - \lambda_1)}\right)$$
$$= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

Our **k-d tree** used the distances generated by this formula to find the four nearest neighbors to each anomaly.

Shortest Distance

We then applied **Lambert's formula for long lines** to the four identified neighbors.

Lambert's formula accounts for the ellipticity of the Earth, which the Haversine formula does not. Consequently, it is accurate on the order of ***10 m***, which is more accurate than the Haversine formula but considerably slower. By using both formulae for successive approximations, we retained the accuracy of Lambert's formula and the speed of the Haversine formula.

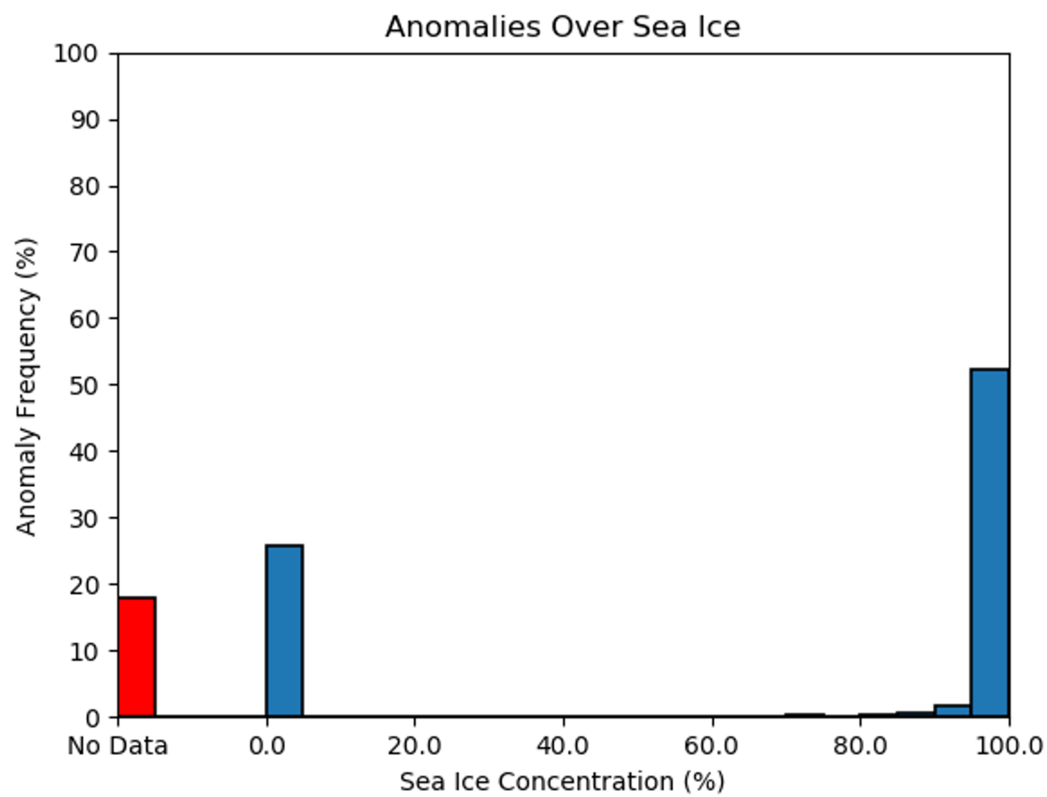
Verifying Overlap

To check whether the resulting single nearest neighbor is collocated with the anomaly, we check whether the distance of the anomaly falls completely within the resolution of the NSIDC AMSR observation.

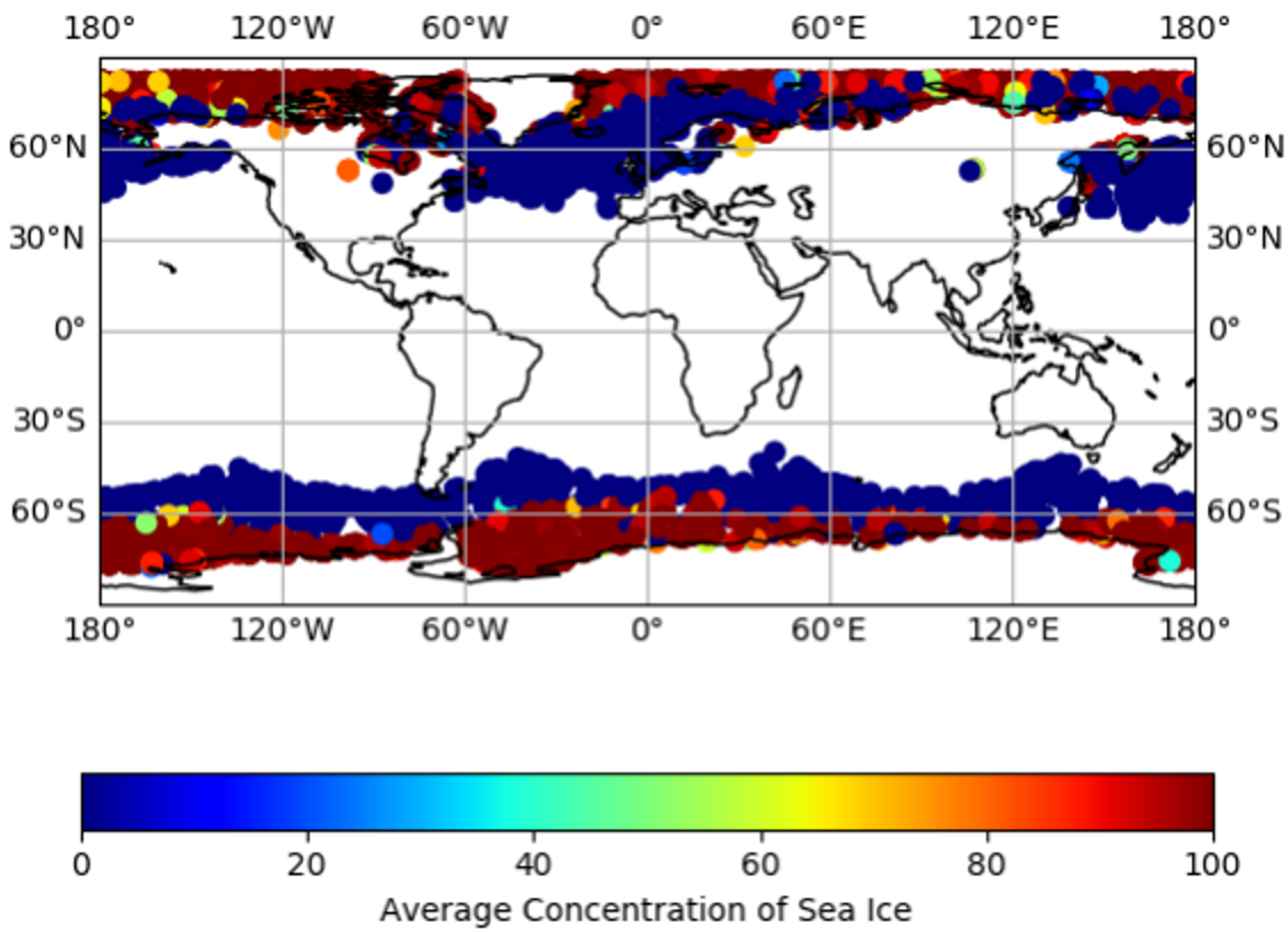
- **Distance ≤ (AMSR resolution / 2) - (CALIOP resolution / 2)**
- **Distance ≤ (25 km / 2) - (1 km / 2)**
- **Distance ≤ 12 km**

Results

The results for the collocation of the 7236 water-phase, single-layer, over-water anomaly cases in 2007 with the sea ice data are presented in the histogram to the right.



Collocated sea ice and anomaly data



Conclusions

- Appears to be a strong correlation between sea ice and the anomalies (~52% over ice for 2007)
- However, sea ice does not account for all of the anomalies and the cause of these other anomalies needs to be understood as well.

3D Radiative Effects Hypothesis

Goal

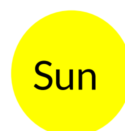
- We want to know where on the cloud the anomaly occurs, specifically if the anomaly is on the illuminated side of the cloud.
- To do this, the slope around the anomaly and the direction of the sun will be determined.
- If the anomaly occurs on a point of the cloud which has a large slope and is in the path of the sun, this would support the hypothesis.

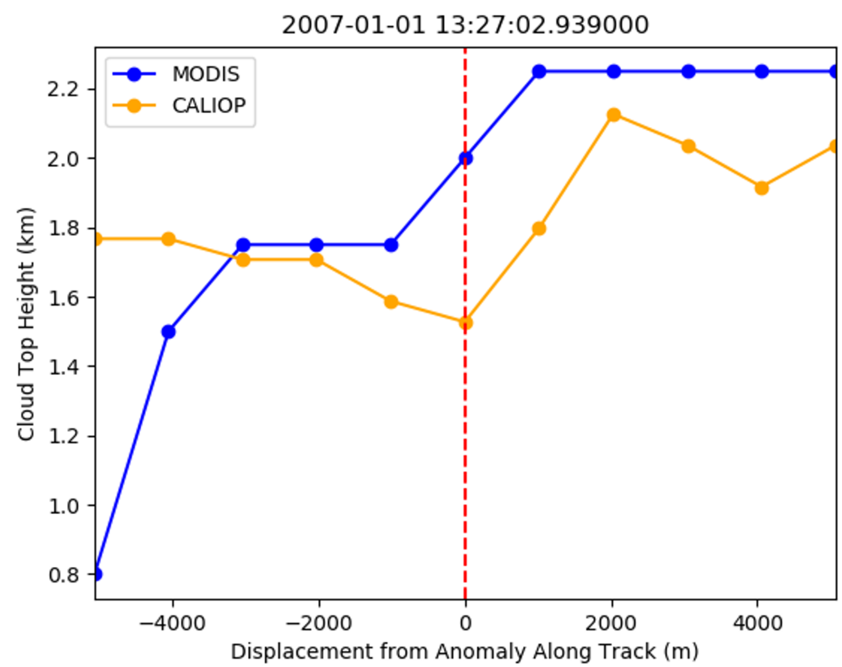
Approach

- The MODIS cloud top height was accessed from the CALIPSO-MODIS collocated data files.
- The CALIOP layer top height was accessed from the CALIPSO data files. The CALIPSO cloud top height was found by including only the layer heights that had their layer identified as a cloud.
- The cloud top height for the anomaly and for the +/- 5 pixels (MOIDS) and lidar shots (CALIPSO) surrounding the anomaly were found and plotted.

Results

- This is an example of the cloud top heights along the track for both MODIS and CALIOP.
- Using the SZA and solar azimuth angle (SAA) from the CALIOP data file and a map showing the direction of the satellites' track, it was determined that the sun was coming from the left of this image, thereby indicating that the anomaly is on the illuminated side of the cloud.



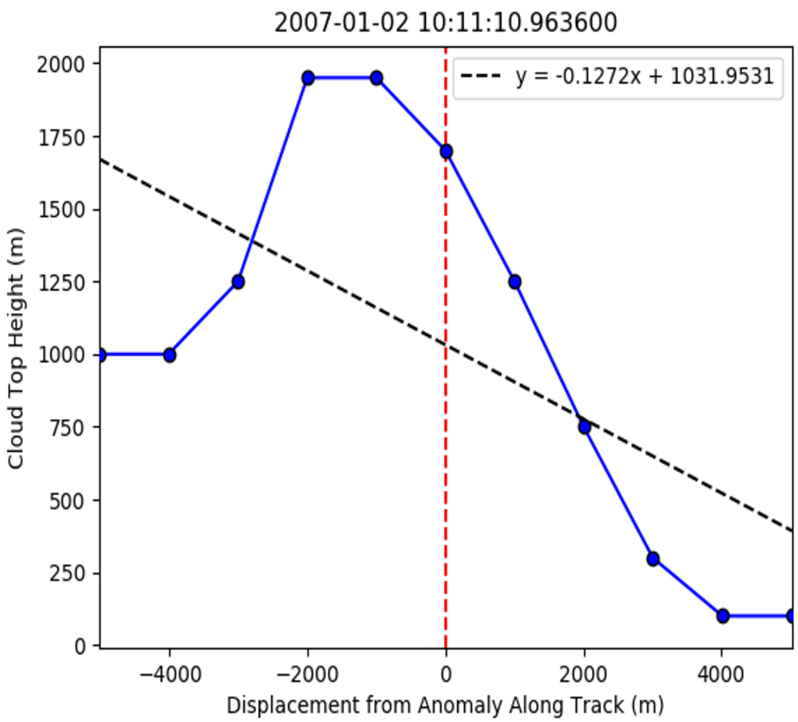


Slope Calculation

- The slope was calculated using least squares linear regression for all 11 MODIS cloud top height points (+/- 5 around the anomaly).
- This was done for anomalies that occurred on *single-layer*, *ice phase clouds in 2007.*

Results

- This is an example of the plotted MODIS cloud top height and corresponding slope (black, dashed line).
- However, the method used to determine the slope is not producing the most accurate representation of the actual steepness of the cloud at the location of the anomaly.



Sun

Summary and Next Steps

Summary

- There appears to be a strong correlation between sea ice and the anomalies, but the correlation does not account for about 50% of the over-water anomalies in 2007.
- It is still possible that 3D radiative effects could be associated with the anomalies, but more work needs to be done to determine if there is a correlation.

Future Work

1. Expand data set for sea ice/snow hypothesis
    a. Beyond 2007
    b. Look into over land anomalies and collocate that with snow data.

Future Work

2. Continue to explore the 3D radiative effects hypothesis
   a. Develop a new method to calculate slope that more accurately represents the slope at the point of the anomaly.
   b. Automate the process of determining if the anomaly is on the illuminated side of the cloud.
      i. Possible method: Use 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.

Future Work (cont.)

3. Searching for and Quantifying Other Relationships
    a. Use machine learning to search for relationships between the anomalies and factors other than sea ice and 3D effects.