2019 Fall Challenge
GLOBE North American Region
TOP OBSERVERS
MacArthur Fundamental School

Reliability of high cloud data between citizen science and GEO satellite during the 2019 Fall Cloud Challenge

Victoria Tlilayatzi 408348@sausdlearns.net

Mia Lacy 407878@sausdlearns.net

Trina Tran 403697@sausdlearns.net

Kimberly Tiempos 406984@sausdlearns.net

Globe Teacher: Esther Flores

MacArthur Fundamental Intermediate School
Santa Ana, Ca

Supervision: Marilé Colón Robles

2020

Pages	Content
3	Abstract
4	Key words
4	Research questions
4	Hypotheses
5	Research Importance
6	Introduction
7	Material and Methods
11	Results and discussions
13	Difficulties
16	Suggestions and recommendations
16	Acknowledgments
17	References

Abstract

The presence of low-level and mid-level clouds can lead to inaccuracies in high cloud data collection. This interference can lead to satellite errors or ground observation errors. Satellite errors can be reduced with changes in instrument sensitivity, while observer errors could be improved with additional training on high cloud data collecting. From October 15 to November 15 of the year 2019, students at Douglas MacArthur Fundamental Intermediate School participated in the Fall Cloud Challenge to gather cloud data to determine how often satellite data and student data agreed. If there was an agreement, it would indicate that low and mid clouds did not interfere with the accuracy of high cloud data collection. The inverse was considered as well. While over 1000 observations were made, a concern arose about the quality of the student observer data. Data was sorted as "reliable" and "unreliable" upon closer analysis. Of the 395 observations considered to be "reliable" data, it was concluded that 328 (83%) agreements were found between student observer high cloud data and satellite high cloud data. At first glance, this did not seem to support our hypothesis. However, it was noted that some of this data was taken in the absence of high clouds or the absence of low/mid clouds which were not ideal for testing the hypothesis. Upon further analysis, in the ideal condition of the presence of both low/mid clouds as well as high clouds, the inverse was found. Out of 82 observations under this ideal condition, only 14 (17%) observations consisted of an agreement of high cloud data between student and satellite. This did support our hypothesis that low/mid cloud presence can interfere with high cloud data recording. Further high cloud data collecting in the presence of both high clouds and low/mid clouds needs to be collected to gather greater support for our hypothesis.

Key words

- High Cloud Cover - percentage of high clouds
- **High Cloud Agreement**- A high cloud match between GEO satellite data and the students data
- Total Cloud Cover - Data used to determine reliability of student data
- **Reliable Data**- Data considered to be of higher quality due to the fact that the satellite and the student Total Cloud Cover data agreed within 20% (ex: If the students reported 15% Total Cloud Cover and the satellite said 25% percent Total Cloud Cover student data was considered "reliable")
- GEO Satellite- The GEO satellite was the prefered satellite for this report due to the amount satellite matches compared to Aqua and Terra satellites that rarely matched
- NASA Globe Observer App- App students used to record cloud data
- **GEO satellite match** - Student cloud observation matched with GEO satellite data based on the time and location of the report providing more information on "reliable" data
- **GLOBE Cloud Satellite comparison**- report showing GEO satellite data and student data in detail used to determine if certain data was reliable.
- Ground Observer Data: Student collected data

Research questions

1. Does the presences of low-level clouds and mid-level clouds affect the accuracy of recording high cloud data?
2. Could we determine the accuracy of high cloud data recording and the effect of low/mid clouds by noting instances in which student high cloud data agreed with satellite high data?
3. How reliable is Observer App data?

Hypotheses

- Low-level and mid-level clouds cause a disagreement between the student High Cloud data and the satellite High Cloud data.
- Distinguishing data as reliable from unreliable, can lead to more accurate reporting of high cloud data agreement and disagreement.

Research Importance

Collecting cloud data can be useful for many scientific studies, including climate change and water availability. A high quality and accurate collection of the cloud data by both ground observers and satellites can ensure better scientific conclusions. And when both agree on the data, this can support the reliability of both types of data. When the inverse occurs, and satellite data does not agree with ground observer data, this could suggest the unreliability of one or both.

A particular concern arose in this report to note disagreement of high cloud data in the presence of low-level clouds and mid-level clouds (Dodson, 2019). In the presence of low/mid clouds, it was expected that more disagreements would occur between satellite data and ground observer data.

In such a case, where the ground observers record unreliable data, it could suggest additional training in data collecting would be needed. One possible method of training would be to allow more time working with a trainer or teacher to ensure accuracy. Better training of teachers may also be needed. In addition, technological tools used to collect data could be improved (Colón Robles, 2019). For example, the GLOBE Observer App has been designed to prompt students to particular questions that better ensure their data collecting is accurate. Further sensitivity of this tool might also be suggested.

The disagreements between satellite data and ground observer data may also reveal a need for adjustments to satellite tool sensitivity. Additional comparisons with multiple observers and/or multiple satellites could help suggest the sensitivity needs of any one satellite. Noting these disagreements could lead to higher quality and accuracy in data collecting.

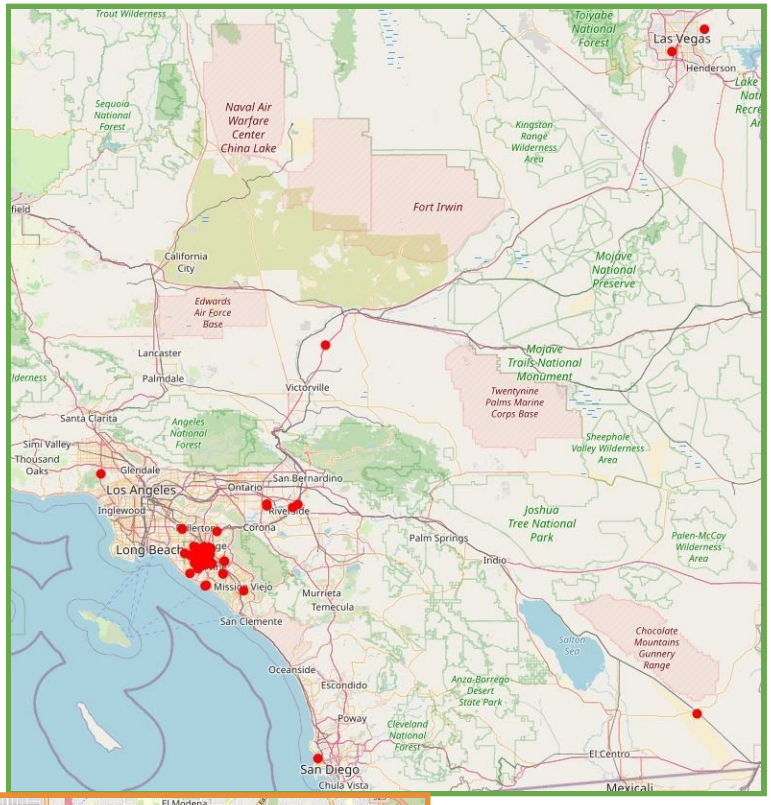
Introduction

From October 15 to November 15 of the year 2019, students at Douglas MacArthur Fundamental Intermediate School participated in the Fall Cloud Challenge. The data was gathered at the convenience of each student and MacArthur was recognized as a Top Observer for the GLOBE North American Region.

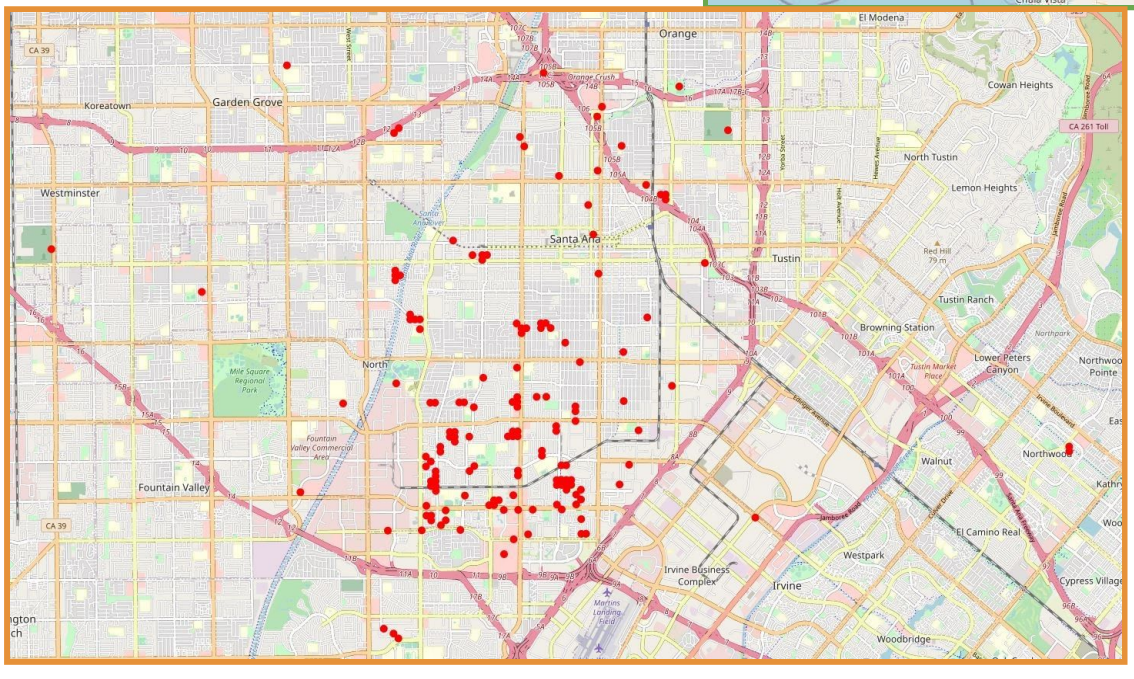
Over 1000 observations were made within that month, not all of them were used in this study. Ground observations were recorded with variability in student, time, and location. The goal was to get students involved in real science data and hopefully gather enough data to analyze various possible topics (Colón Robles, 2019).

In analyzing this data, many challenges arose that lead to a need to separate data into two categories: reliable and unreliable. The purpose was to ensure the data used in further analysis was of high quality.

Map of Observation Locations



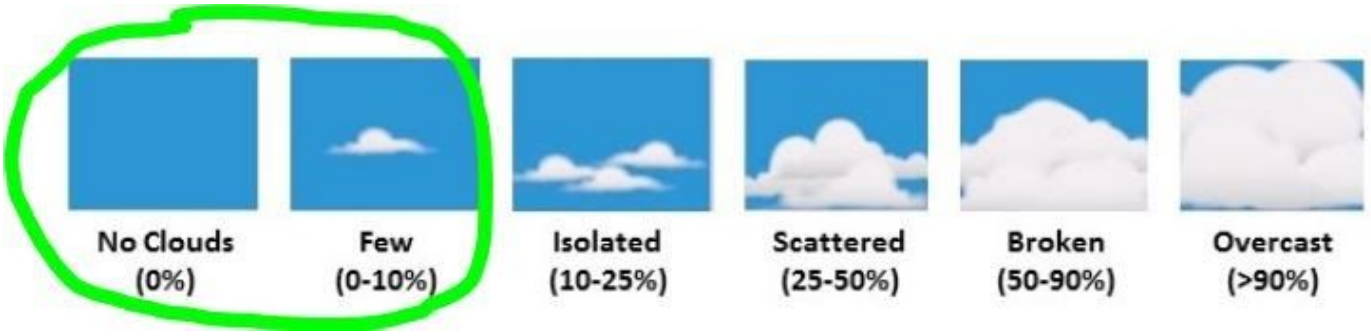
Map of Orange County Observation Locations



Material and Methods

- **NASA Globe Observer App**
- **phone or digital device with a camera**
- **Cloud Chart**

1. Cloud data could be recorded on paper or directly inputted into the NASA Globe Observer App on a phone or other digital device with a camera.
2. All Globe Observer App data was compiled in a Google Sheet for analysis.
3. Data was categorized as reliable and unreliable based on criteria discussed below.
4. Percentages for high cloud cover were taken from NASA Globe website (The categories for No Cloud and Few Clouds (0-10%) were merged:



5. Data was categorized as agreeing or disagreeing between ground observer high cloud data and GEO satellite high cloud data to investigate hypotheses.

Reliable and unreliable data

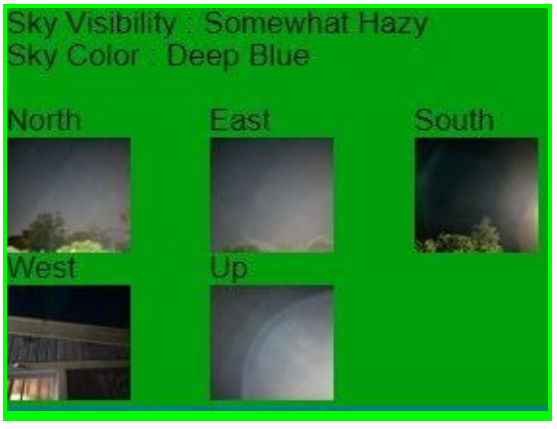
1. Initially, over 1000 ground observations were recorded on either paper or through the Globe Observer App.
2. To increase data quality, only data taken with Globe Observer App was further analyzed since the reliability of this data increases with the App guiding cloud accuracy. The Globe Observer App totaled 581 observations.
3. Of the 581, only 511 were GEO satellite matches that provided a **GLOBE Cloud Satellite comparison**; the rest were considered unreliable.

4. Of the 511, it was determined that a disagreement in total cloud cover between the GEO satellite and ground observation would represent too great a difference to consider reliable. Reliable data was determined to be that which had no greater difference in total cloud cover than 20% between satellite and observer. That left 395 observations.

5. Questionable data was individually analyzed and determined to be reliable or unreliable. The following table shows examples. In order to sort the reliable and unreliable data, we had to analyze the **GLOBE Cloud Satellite comparisons**.

Satellite		**GEO**	**Your Observation**
Universal Date/Time 2019-11-12		04:35	04:24
Latitude Range Longitude Range		33.38 to 34.02 -118.23 to -117.59	**Latitude 33.69** **Longitude -117.9**
Total Cloud Cover		Isolated 19.19%	**Scattered (25-50%)**
HIGH	Cloud Cover Cloud Altitude Cloud Phase Cloud Opacity	No Clouds	Contrails: Spreading 1 Cirrocumulus Scattered (25-50%) Transparent
MID	Cloud Cover Cloud Altitude Cloud Phase Cloud Opacity	No Clouds	
LOW	Cloud Cover Cloud Altitude Cloud Phase Cloud Opacity	Isolated 19.19% 0.82 (km) Water 284.89 (K) Transparent	
Corresponding NASA Satellite Images. Click to view image --->		***GOES-15*** Visible Infrared 	Sky Visibility : Clear Sky Color : Blue

6. Data that might be unreliable was analyzed individually to see if it provided enough information to be useful in observing high cloud data. The following are some examples:

Reliable or Unreliable Examples	Photo
Some students experienced obstructed views that made it difficult to accurately collect data. Reliable because evidence for type of cloud and percentage can still be gathered from the photo.	
Some students confused the sky obscured for cloud cover. Reliable because the photos taken can still provide data on high clouds.	
Student agreed with the satellite, but photos show no clouds. Unreliable because photos do not show what the student reported.	
Students collected data at night, when it was difficult to determine clouds. Unreliable due to the difficulty in knowing whether the disagreement or agreement of high cloud data was visible.	

7. Of the 395 reliable observations, data was further separated into the following categories:
    a. Total high cloud agreement between satellite and observer
    b. Total high cloud agreement in the presence of high clouds
    c. Total high cloud agreement in the absence of high clouds
    d. Total high cloud agreement in the presence of low/mid clouds
    e. Total high cloud agreement in the presence of high clouds and low/mid clouds

Results and discussions

Douglas MacArthur Fundamental Intermediate School participated in the 2019 Fall Cloud Challenge to gather cloud data. While over 1000 observations were made, a concern arose about the quality of the ground (student) observer data. Data was sorted as reliable and unreliable upon closer analysis. The reliable data was further analyzed to determine whether the presence of low-level and mid-level clouds caused inaccuracies that lead to disagreement between satellite high cloud data and ground observer high cloud data as stated in our hypothesis.

Of the over 1000 observations, 581 ground observations were recorded using Globe Observer App, which was considered to control for higher quality data collecting and was considered more reliable (**Figure 1**). Of the 581 ground observations, 511 were GEO satellite matches. This meant that the GEO satellite and ground observation was recorded within 15 minutes of each other and therefore could reveal agreement and disagreement in high cloud data. Of the 511 GEO satellite matches, only 395 were considered reliable enough to use for further analysis of the agreement and disagreement of high cloud data (**Table 1**). To determine ground observations were reliable, the total high cloud cover needed to be within 20% variation of the satellite data. High cloud cover that was greater than 20% different, was considered unreliable. In addition, reliability was determined by analyzing photos and **GLOBE Cloud Satellite comparison** charts.

Figure 1. Determining reliable data

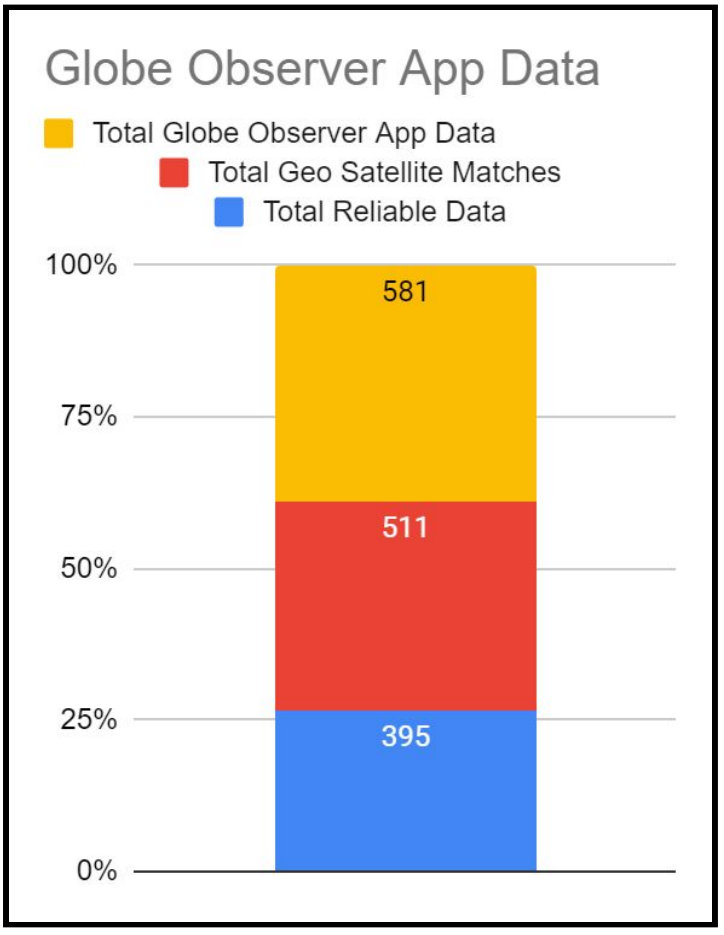


Table 1. Chart of reliable data

Total Globe Observer App Data	Total Geo Satellite Matches	Total Reliable Data
581	511	395

The 395 observations that were considered reliable, were further analyzed and categorized (**Table 2**). Of the 328 instances of agreement between the satellite and ground observer (student), there was agreement in 83% of the reliable data. This appeared to not support our hypothesis.

Table 2. Reliable data categories

Analysis of Reliable Data = within 395 Observations			
High Cloud Data Agreement between satellite and student	**High Cloud Agreement in High Cloud presence**	**High Cloud Agreement collected in High Cloud absence**	**Data collected in Presence of low/mid clouds**
328	82	246	178
83%	20.8%	62.3%	45.1%

However, it was also noted that some of the above categories show a lack of essential components that were expected to draw conclusions in our hypothesis. For example, some of the observations were recorded without the presence of high clouds. Other data was collected without the presence of low/mid clouds. Having both high clouds and low/mid clouds present would lead to an ideal condition to investigate the interference of low/mid clouds on data collection of high clouds.

Further analysis was needed in order to determine the results in the ideal condition. By ideal condition, it was data collected in the presence of both high clouds and low/mid clouds. This condition would best show the accuracy of agreement between the satellite and ground observation on high cloud data. However, this reduced the amount of data drastically. Under these conditions, only 82 observations were left (**Table 3**). Within these 82, only 14 or 17% if the total observations contained agreeing high cloud data.

Table 3. Reliable data under ideal condition

Analysis of data in presence of high and low/mid clouds = 82		
All data collected in ideal condition	**Total High Cloud Agreement Presence of low/mid clouds**	**Total High Cloud Disagreement in presence of low/mid clouds**
82	14	68
100%	17%	83%

Conclusion

While this supports our initial hypothesis that low/mid clouds interfere, and therefore cause more disagreement, between ground observations and satellites, the available data was determined not to be enough to fully support our hypothesis. More data, with ideal conditions, would need to be collected.

Difficulties

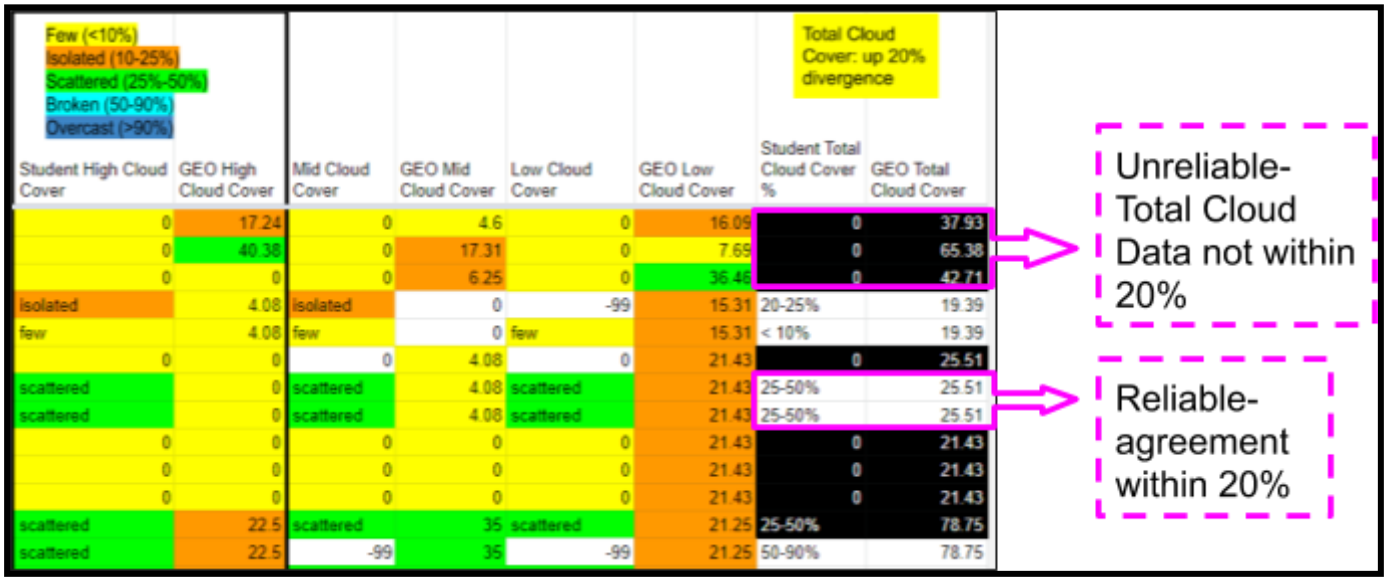
To handle the large amount of data, the team utilized Google Sheet. This allowed for team members to individually analyze different categories of data. Data was color coded to help spot high cloud agreement/disagreement and other categories.

Figure 2. Google sheet of categories

High Clouds Present	High Cloud Data Agrees	Student High Cloud Cover	GEO High Cloud Cover	Low or Mid Clouds Present	Mid Cloud Cover	GEO Mid Cloud Cover	Low Cloud Cover	GEO Low Cloud Cover	GEO/Terra/Aqua Satellite Match Table	Student Total Cloud Cover %	GEO Total Cloud Cover
x		isolated	4.08	x	isolated	0	-99	15.31	https://scool.la	20-25%	19.39
x		isolated	5.21	x	few	9.38	isolated	11.46	https://scool.la	20-25%	26.05
x		isolated	8.6	x	-99	5.38	-99	23.66	https://scool.la	25-50%	37.64
x		isolated	21.43	x	-99	4.08	isolated	6.12	https://scool.la	20-25%	31.63
x		isolated	21.43	x	-99	4.08	-99	6.12	https://scool.la	25-50%	31.63
x	x	isolated	22.55	x	-99	1.96	-99	31.37	https://scool.la	20-25%	55.88
x		isolated	50.6	x	scattered	21.69	broken	2.41	https://scool.la	50-90%	74.7
x		isolated	90.22	x	-99	9.78	-99	0	https://scool.la	20-25%	100
x		overcast	0	x	overcast	0	overcast	16.33	https://scool.la	90-100%	16.33
		overcast	0	x	overcast	11.34	overcast	42.27	https://scool.la	90-100%	53.61
		overcast	0	x	-99	22.11	-99	43.16	https://scool.la	90-100%	65.27
		overcast	0	x	overcast	42.55	overcast	54.26	https://scool.la	90-100%	96.81
x		overcast	0	x	overcast	0	overcast	9.18	https://scool.la	90-100%	9.18
x		overcast	0	x	overcast	0	overcast	22.92	https://scool.la	90-100%	22.92
x		overcast	0	x	overcast	0	overcast	79.59	https://scool.la	90-100%	79.59
x		overcast	4.17	x	-99	1.04	-99	62.5	https://scool.la	90-100%	67.71
x		overcast	5.43	x	overcast	13.04	overcast	11.96	https://scool.la	90-100%	30.43
x		overcast	9.28	x	overcast	84.54	overcast	6.19	https://scool.la	90-100%	100.01
x	x	overcast	73.91	x	overcast	7.61	overcast	1.09	https://scool.la	90-100%	82.61
x		overcast	85.42	x	overcast	14.58	overcast	0	https://scool.la	90-100%	100
x	x	overcast	93.41	x	-99	3.3	-99	0	https://scool.la	90-100%	96.71
x	x	overcast	100	x	overcast	0	overcast	0	https://scool.la	90-100%	100
x	x	overcast	100	x	overcast	0	broken	0	https://scool.la	90-100%	100
x	x	overcast	100	x	overcast	0	overcast	0	https://scool.la	90-100%	100
x		scattered	0	x	scattered	4.08	scattered	21.43	https://scool.la	25-50%	25.51
x		scattered	0	x	scattered	4.08	scattered	21.43	https://scool.la	25-50%	25.51
x		scattered	0	x	scattered	0	scattered	42.27	https://scool.la	25-50%	42.27
x		scattered	0	x	scattered	0	isolated	42.27	https://scool.la	25-50%	42.27
x		scattered	0	x	broken	0	-99	34.34	https://scool.la	50-90%	34.34
x		scattered	0	x	scattered	0	isolated	10.1	https://scool.la	25-50%	10.1
x		scattered	0	x	-99	6.52	scattered	21.74	https://scool.la	25-50%	28.26
x		scattered	0	x	-99	0	-99	17.89	https://scool.la	25-50%	17.89
		scattered	0	x	scattered	0	scattered	22.99	https://scool.la	25-50%	22.99
		scattered	0	x	-99	7.22	scattered	47.42	https://scool.la	25-50%	54.64

Even though the data appeared to be organized, we had to allow a certain amount of human error. This led to our categorizing data as reliable or unreliable. This task required identifying when the ground observer's total cloud cover data agreed within 20% of the satellite data.

Figure 3. Reliable and unreliable data categorized



Few (<10%)
Isolated (10-25%)
Scattered (25%-50%)
Broken (50-90%)
Overcast (>90%)

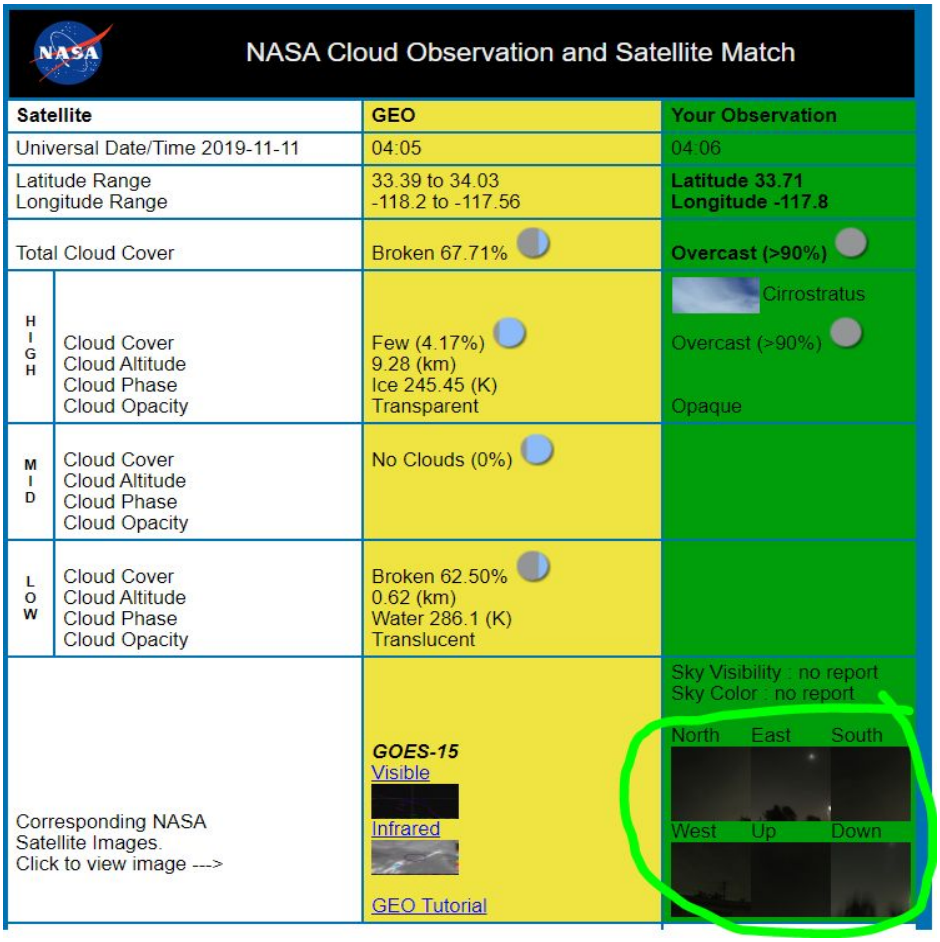
Total Cloud Cover: up 20% divergence

Student High Cloud Cover	GEO High Cloud Cover	Mid Cloud Cover	GEO Mid Cloud Cover	Low Cloud Cover	GEO Low Cloud Cover	Student Total Cloud Cover %	GEO Total Cloud Cover
0	17.24	0	4.6	0	16.09	0	37.93
0	40.38	0	17.31	0	7.69	0	65.38
0	0	0	6.25	0	36.46	0	42.71
isolated	4.08	isolated	0	-99	15.31	20-25%	19.39
few	4.08	few	0	few	15.31	< 10%	19.39
0	0	0	4.08	0	21.43	0	25.51
scattered	0	scattered	4.08	scattered	21.43	25-50%	25.51
scattered	0	scattered	4.08	scattered	21.43	25-50%	25.51
0	0	0	0	0	21.43	0	21.43
0	0	0	0	0	21.43	0	21.43
0	0	0	0	0	21.43	0	21.43
scattered	22.5	scattered	35	scattered	21.25	25-50%	78.75
scattered	22.5	-99	35	-99	21.25	50-90%	78.75

Unreliable- Total Cloud Data not within 20%

Reliable- agreement within 20%

Figure 4. GLOBE Cloud Satellite comparison



NASA Cloud Observation and Satellite Match

Satellite		GEO	Your Observation
Universal Date/Time 2019-11-11		04:05	04:06
Latitude Range Longitude Range		33.39 to 34.03 -118.2 to -117.56	**Latitude 33.71** **Longitude -117.8**
Total Cloud Cover		Broken 67.71%	**Overcast (>90%)**
HIGH	Cloud Cover Cloud Altitude Cloud Phase Cloud Opacity	Few (4.17%) 9.28 (km) Ice 245.45 (K) Transparent	Cirrostratus Overcast (>90%) Opaque
MID	Cloud Cover Cloud Altitude Cloud Phase Cloud Opacity	No Clouds (0%)	
LOW	Cloud Cover Cloud Altitude Cloud Phase Cloud Opacity	Broken 62.50% 0.62 (km) Water 286.1 (K) Translucent	
Corresponding NASA Satellite Images. Click to view image --->		***GOES-15*** Visible Infrared GEO Tutorial	Sky Visibility : no report Sky Color : no report North East South West Up Down

While it initially appeared to be a manageable task, we identified some data that was not easily categorized. For instance, some ground data showed variation between the photographs taken and the data recorded because the photo was taken at night (**Figure 4**). The team had to individually determine the reliability of some ground observations by analyzing the **GLOBE Cloud Satellite comparison** charts.

Figure 5. Data containing "-99"

Satellite		GEO	Your Observation
Universal Date/Time 2019-11-12		04:35	04:24
Latitude Range Longitude Range		33.38 to 34.02 -118.23 to -117.59	**Latitude 33.69** **Longitude -117.9**
Total Cloud Cover		Isolated 19.19%	**Scattered (25-50%)**
HIGH	Cloud Cover Cloud Altitude Cloud Phase Cloud Opacity	No Clouds	Contrails: Spreading 1 Cirrocumulus Scattered (25-50%) Transparent
MID	Cloud Cover Cloud Altitude Cloud Phase Cloud Opacity	No Clouds	
LOW	Cloud Cover Cloud Altitude Cloud Phase Cloud Opacity	Isolated 19.19% 0.82 (km) Water 284.89 (K) Transparent	
Corresponding NASA Satellite Images. Click to view image --->		***GOES-15*** Visible Infrared 	Sky Visibility : Clear Sky Color : Blue

Some data contained "-99" which usually meant that information was not provided by the ground observer. At times, this indicated that the observer meant to record that there were no clouds in a category, but skipped the questions instead. In certain cases, the "-99" required further analysis and discussion before determining its reliability (**Figure 5**).

Suggestions and recommendations

1. More data needs to be collected that fulfills that ideal conditions of high cloud data collection in the presence of low-level and mid-level clouds.
2. Further analysis is needed to interpret the 2 distinct results from our high cloud data agreements. When observing all high cloud data agreement, we found an 83% agreement between ground observations and satellite data on high clouds. However, in instances where both high clouds and low/mid clouds were present, only 14% agreement was found.
3. Continued studies on categorizing data for reliability and creating standards for reliability could help guide this research.
4. Continued studies can determine whether the disagreement is due to a need for more sensitivity of satellite instruments or additional training of participants in ground observations.
5. Further strategies to inform and train ground observers could improve data reliability. These strategies could be tested to find their impact of gathering reliable data.

Acknowledgments

Our team would like to thank Mr. Peter Falcon for providing essential guidance that helped initiate our data collecting and investigation. Also, for providing us with resources and information that supported us throughout our cloud data research, including training our teacher, Mrs. Flores, so that she could help us. We would also like to thank the many individuals that make Globe Student Research Symposium possible, since life-long skills and experiences are acquired through this opportunity. Special thanks to Jennifer Bourgeault for informing and updating our team through the stay-at-home orders. We would also like to thank Mr. John Power, Ms. Walter, and Ms. Patter from UCI for asking us important questions about our data, and helping us look at our data through different perspectives. Our team would also like to acknowledge Ms. Valerie Armstrong, MacArthur's 6th grade science teacher, for her input on our data analysis, and for inviting UCI graduate students to discuss our data. Finally, we cannot imagine the completion of this research without the time and guidance of Mrs. Marilé Colón Robles. We are grateful for her support and encouragement throughout this entire time, without which we may have given up or may have become lost in cloud data.

References

Atkinson, Joe. "GLOBE Fall Cloud Challenge Rakes in the Observations." *NASA*, NASA, 22 Nov. 2019, www.nasa.gov/feature/langley/globe-fall-cloud-challenge-rakes-in-the-observations

Dodson, J. Brant, et al. "Eclipse across America: Citizen Science Observations of the 21 August 2017 Total Solar Eclipse." *Journal of Applied Meteorology and Climatology*, vol. 58, no. 11, 2019, pp. 2363–2385., doi:10.1175/jamc-d-18-0297.1.

Dunbar, Brian. "Why Does NASA Have Its Head in the Clouds?" *NASA*, NASA, www.nasa.gov/audience/foreducators/5-8/features/F_Head_in_the_Clouds.html.

"GLOBE Observer." *GLOBE.gov*, observer.globe.gov/do-globe-observer/clouds.

"NASA GLOBE Cloud Data." *NASA*, NASA, scool.larc.nasa.gov/cgi-bin/NASA-GLOBESatMatch.cgi?observation_id=116-173847-3226 3218-201911110406.

Colon Robles, Marile, et al. "Integrating Tech: Making Science Come Alive With Clouds." *Science Scope*, vol. 043, no. 04, 2019, doi:10.2505/4/ss19_043_04_8.

Colón Robles, Marilé, et al. "Clouds Around the World: How a Simple Citizen Science Data Challenge Became a Worldwide Success." *Bulletin of the American Meteorological Society*, 2020, doi:10.1175/bams-d-19-0295.1.