Do citizen science Intense Observation Periods increase data usability? A deep dive of the NASA GLOBE Clouds data set with satellite comparisons

J. Brant Dodson[1], Marilé Colón Robles[1], Tina M. Rogerson[1], and Jessica E. Taylor[2]

[1] Science Systems and Applications, Inc., Hampton, VA 23666, USA.

[2] NASA Langley Research Center, Lidar Science Branch, Hampton, VA 23681-2199, USA.

Corresponding author: J Brant Dodson (jason.b.dodson@nasa.gov)

Key Points:

- NASA GLOBE Clouds citizen science reports of cloud cover by level agree closely with co-located satellite observations.
- Disagreements about low clouds relate to high clouds blocking the satellites' view, but disagreements about high clouds relate to opacity.
- Through carefully designed methodology and testing for robustness, citizen science can provide a useful data source for scientific research.

Abstract

The Global Learning and Observations to Benefit the Environment (GLOBE) citizen science program has recently conducted a series of month-long intensive observation periods (IOPs), asking the public to submit daily reports on cloud and sky conditions from all regions of Earth. This provides a wealth of crowdsourced observations from the ground, which complements other conventional scientific cloud data. In addition, the GLOBE reports are matched in space and time with geostationary and low Earth orbit satellites, which allows for a straightforward comparison of cloud properties, and minimizes the biases associated with mismatched sampling between participants and satellites.

The matched GLOBE dataset is used to calculate the mean observed cloud cover by atmospheric level both worldwide and by region. The overall magnitudes of cloud cover between the GLOBE participants and the matched satellites agree within 10%, which is notable given the distinctly different natures of the data sources. The mean vertical cloud profiles show GLOBE reporting more low-level clouds and fewer high-level clouds than satellites. The low cloud disagreement is likely related to satellites missing low clouds when high clouds block their view. Conversely, the high cloud disagreement is related primarily to cloud opacity, as satellites may miss some optically thin clouds. Monte Carlo testing shows the results to be robust, and the tripled amount of IOP data reduces uncertainty by half. These findings also highlight ways in which citizen science IOP data may be used to support scientific research while accounting for their unique properties.

Plain Language Summary

Citizen science is becoming an increasingly prominent aspect of scientific research, and so it is important to study how citizen science data can be used effectively. For example, The GLOBE Program has recently conducted a series of special data-collecting events, or “challenges”, which gathered large numbers of reports on cloud and sky conditions. Because NASA GLOBE Clouds matches the participant reports with cloud observations from satellites, we can use these data to get a combined view of clouds from above and below. When looking at the average cloud cover for different atmospheric levels across Earth, we find that the GLOBE participants and the satellites agree quite closely. This is a surprising and fascinating find, given how different in nature volunteer ground reports are to satellite measurements. However, there are some small but notable disagreements between GLOBE participants and satellites about the distribution of cloud cover at different levels. In addition, by testing the data for uncertainty, we show that the results from the GLOBE data are reliable, and that more public participation improves the reliability. So, by carefully designing the analysis methodology, and by testing for the uncertainty of the data, citizen science can make a meaningful contribution to scientific research.

1 Introduction

Citizen science is increasingly becoming an important component of the scientific endeavor. Data from public participants are collected and used for a variety of scientific fields, with atmospheric science being just one among many. The application of citizen science data to answering scientific questions about weather and climate has previously been demonstrated for major geophysical events, specifically solar eclipses. The meteorological effects of multiple eclipses have been observed by citizen scientists (Barnard et al., 2016; Hanna et al., 2016; Portas

et al., 2016; Hanna, 2018, Rahman et al., 2019), and recently the Global Learning and Observations to Benefit the Environment (GLOBE) Program had a major campaign to gather meteorological observations of the 21 August 2017 total solar eclipse crossing North America called *How Cool is the Eclipse* (GLOBE, 2021a). These data were useful in quantifying the transient temperature depression associated with the eclipse, and the effect of cloud cover on the amplitude of the depression (Dodson et al. 2019).

Beyond solar eclipses, citizen science data have been useful for examining other major meteorological events (e.g. Horton et al., 2021). In retrospect it seems obvious that major events like these would attract the interest of citizen scientists, and in turn their reports would be useful for better understanding these events scientifically. But this raises the question of using citizen science data for situations where there is no major event to attract public attention. Can citizen science observations be used to address important scientific questions about weather and climate outside of the extremes?

Answering this question involves addressing two issues. First is collecting a sufficient amount of data to generate statistically robust results. Outside of major geophysical events, citizen science projects have the option of soliciting frequent observations during a specific period of time. The GLOBE Program has called these month-long events as “data challenges”, enticing the public to contribute observations gathering large numbers of reports (Colón Robles et al., 2020). These events are similar to the intensive observation periods (IOPs) frequently conducted by long-term research field campaigns, and so we use the terminology in this paper. These IOPs involve conducting publicity campaigns to raise public awareness and enthusiasm of the citizen science project, and thus encourage greater participation. For example, GLOBE has recently conducted three IOPs, the Spring Cloud Challenge in 2018 (hereafter known as SCC18), the Fall Cloud Challenge of 2019 (FCC19), and the Community Cloud Challenge of 2020 (CCC20). These IOPs boosted the collection rate of cloud reports considerably above normal, and provide a larger collection of data which is useful for addressing scientific research questions. The success of these Challenges has encouraged other GLOBE projects to develop their own IOPs to boost data collection, such as for land cover (Kohl et al., 2021). Such IOPs may become a more important aspect of citizen science efforts in the future.

The second issue is identifying scientific questions that citizen science data are particularly useful for addressing. One such question in atmospheric science involves precise measurements of the vertical structure of cloud cover, which has important implications for weather and climate such as future climate change projections (e.g. Wielicki et. al, 1995; Wang et al., 1998; Barker et al., 1999; Wang et al., 2021), but is difficult to observe with satellites in some situations. In particular, satellites with passive visible and infrared sensors have difficulty observing clouds in certain situations, such as low clouds overlapped with high clouds (e.g. Wielicki et. al, 1995; Chang and Li, 2005), and optically thin high clouds. This limitation has led to various assumptions being employed in the creation of cloud climatologies. More recently, satellites with active sensors, such as CALIPSO and CloudSat, have been used to improve observations of vertical cloud structure (e.g., Sassen et al., 2008; Kato et al., 2010; Li et al., 2015; Oreopoulos et al., 2017; L’Ecuyer et al., 2019; Hang et al., 2019). While these tools have proven useful, they still have limitations, such as the limited sampling of the diurnal cycle, that provide opportunities for additional investigation.

Citizen science cloud observations from GLOBE provide a potentially useful source of information to further improve knowledge of vertical cloud structure. A primary benefit of the

GLOBE dataset is that the participant reports are automatically matched with observations from both geostationary and polar-orbiting satellites carrying passive visible/infrared sensors. These satellites observe large regions of Earth many times a day and can be matched with participant reports more regularly than most other cloud-observing systems (both spaceborne and ground-based). The large volume of pre-matched data allow for a detailed comparison of the vertical cloud structure, and avoid biases in mean cloud properties caused by differing sampling statistics between ground reports and satellites, which otherwise is problematic (e.g. Sassen et al., 2008). And the inclusion of multiple satellites from two different orbits in the matched data increases the robustness of the comparison beyond that of using a single satellite.

An example of a GLOBE report matched with both GEO and CALIPSO observations may help illustrate the utility of the GLOBE reports. Figure 1 shows the satellite match tables for a GLOBE report from 11 November 2019 near Toulouse, France. The GLOBE participant reported broken mid-level clouds, and scattered low-level clouds, with clouds at both levels being opaque, and no high-level clouds reported. The report also included photographs, which appear to corroborate the reported cloud coverage. Near the same time, METEOSAT-11 detected 100% opaque high cloud cover, with no detection of mid- and low-level clouds. The attached visible and infrared images (Fig. 2a and 2b) depict overcast conditions with cloud top temperatures near -30°C to -40°C, consistent with a high cloud layer.

While these reports may be seemingly difficult to reconcile at first, this particular report was also one of a few dozen that was also co-located with a CALIPSO overpass (Fig. 2c). CALIPSO matches are too rare for use in deriving robust statistical results, but they can be used for case studies like this. The CALIPSO swath depicts a high cloud layer, with cloud tops about 8 km altitude, over and near the participant's location. Surrounding the high cloud layer, to the north and south, is a low cloud layer with cloud tops at about 2 km altitude. Most of the region under the high cloud layer is hidden from CALIPSO's view, but there are a few narrow gaps in the high cloud layer that reveal mid-level clouds with heights of 3-5 km. These clouds are typically completely hidden from the geostationary satellite but are in clear view of the GLOBE participant. Thus, the participant and the satellites provide complementary views of the same sky conditions.

This example is one of many that show the potential utility of GLOBE data in identifying issues in the satellite data record and providing an additional basis from which to build a correction methodology. It is a continuation of earlier efforts using the predecessor of NASA GLOBE Clouds, the Students' Cloud Observation Online (S'COOL) (Chambers et al., 2003; Chambers et al., 2017). The path for continued progress is aided by the large number of observations collected in the GLOBE IOPs.

The main scientific topic this paper will address is a comparison of global and regional cloud cover as reported by GLOBE citizen scientist participants and by the matched satellite cloud retrievals, with an emphasis on vertical cloud structure. To accomplish this, there will be three specific questions that we will address. First, how closely does the vertical cloud structure estimated from satellite data compare with that from GLOBE data? Second, how robust are the results from the GLOBE data? Finally, if the disagreements in vertical structure between GLOBE and the satellites are not merely noise in the GLOBE reports, then what are the possible reasons for those disagreements?

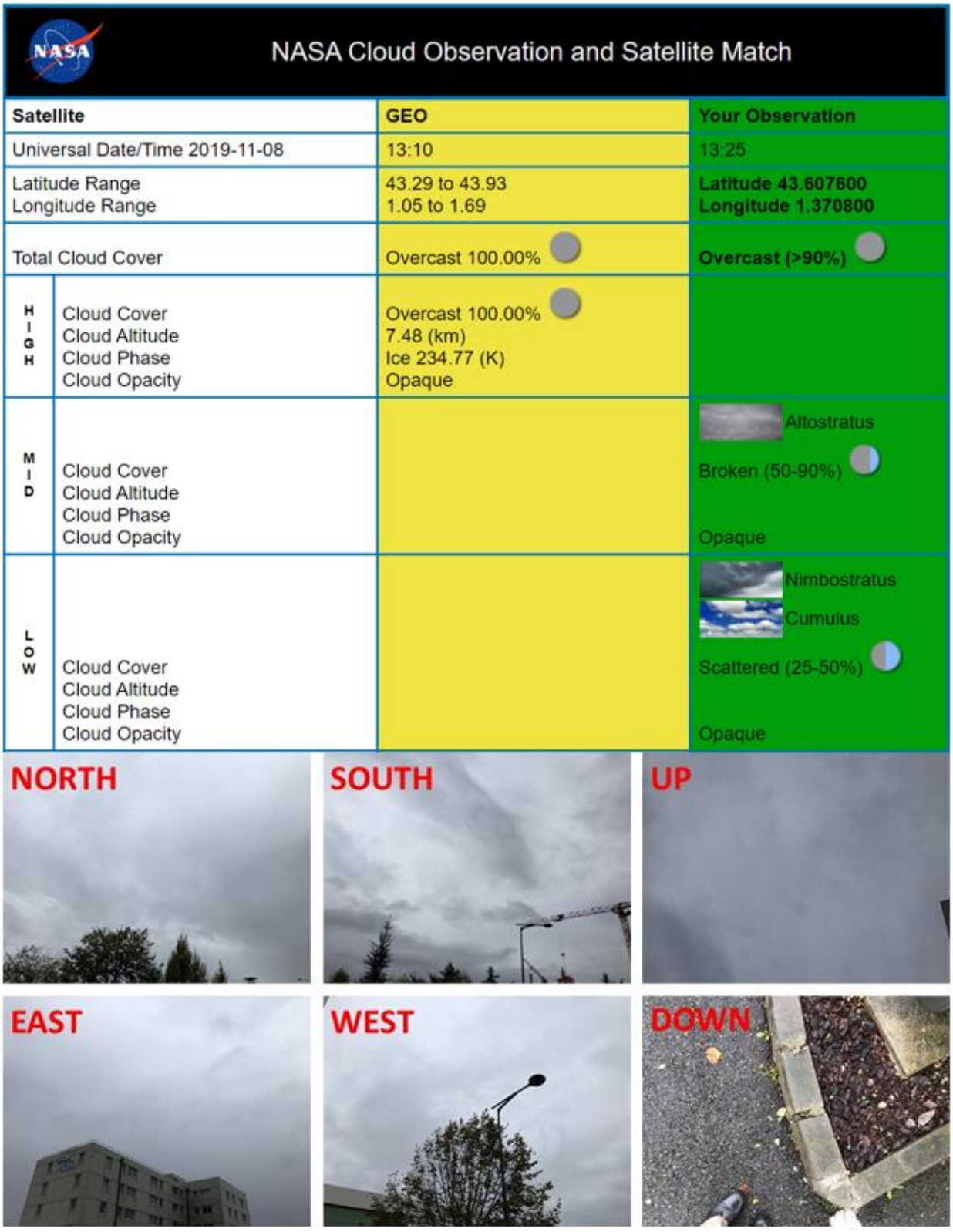

Figure 1. Satellite match table and associated photographs of observation #914394 taken on 11 November 2019 near Toulouse, France, where the GLOBE participant report has been successfully matched with a METEOSAT-11 observation (GLOBE 2022a). In the table, the

left column (in white) lists all the data types that have been matched between the participant and satellite, including time, location, total sky conditions, and cloud properties by altitude. The middle (yellow) column displays the data from METEOSAT-11, which detected only a high opaque ice cloud layer completely covering the area. The right (green) column shows the GLOBE participant report, which does not include the presence of high clouds, but instead middle and low clouds. The photos were taken automatically by the GLOBE Observer app by aiming the device in the six indicated directions.

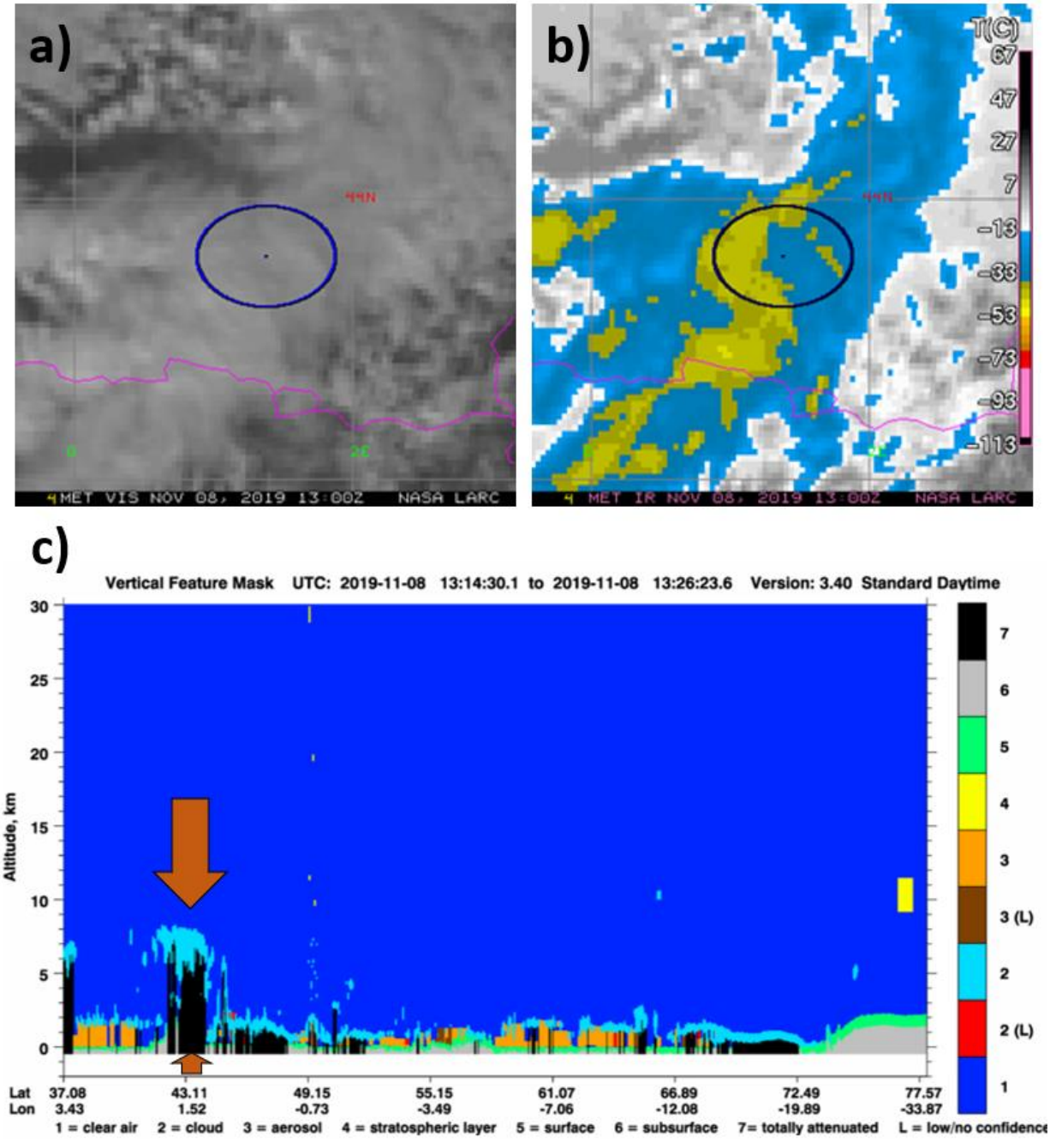

Figure 2. (a) Visible image taken by METEOSAT-11, matched with observation #914394. The central dot indicates the location of the GLOBE participant, while the oval indicates

the 40 km radius around the participant from which the satellite cloud properties are calculated for the satellite match table (shown in Fig. 1). (b) Infrared image from METEOSAT-11. (c) CALIPSO Vertical Feature Mask data (GLOBE 2022b). The x-axis displays the latitude and longitude of each detected feature, which corresponds closely with the satellite location during the overpass, as CALIPSO observes near-nadir. The y-axis displays the altitude of each feature, effectively providing a cross-section view of the atmosphere. The colors correspond with classification of the detected features, with each classification type assigned a number listed at the bottom. Note that cloud cover often hides lower-level features, preventing detection by CALIPSO. The approximate location of the GLOBE participant (43.61°N, 1.37°E) is indicated on the x-axis by the orange arrows.

Section 2 describes the GLOBE data, the SCC18 and FCC19 events, the satellite matching process and resulting data, and the corrections to the dataset needed for a more robust analysis and set of results. Section 3 describes the global and regional comparisons. Section 3 also shows the comparison of cloud cover between the GLOBE participants and the satellite data, including an interesting relationship between the vertical profiles derived from each data source. Section 4 uses a Monte Carlo random sampling technique to test the robustness of the results from Section 3 and quantify the noise of GLOBE data. Section 5 investigates the discrepancies between GLOBE and satellite results more closely, and determines different reasons for disagreements between low cloud cover and high cloud cover. Finally, Section 6 summarizes the results and describes the implication for both estimating the vertical cloud profile and for future studies using GLOBE data.

2 Data and Methods

The GLOBE Program is an international science and education program that launched in 1995 (globe.gov) and provides students and citizen scientists the opportunity to contribute to Earth observations (Berglund, 1999; Finarelli, 1998; Means, 1998; Muller et al., 2015; Nugent, 2018). The GLOBE Program is sponsored by NASA and is supported by the U.S. Department of State, the National Science Foundation, and the National Oceanic and Atmospheric Administration (NOAA). The University Corporation for Atmospheric Research (UCAR) manages The GLOBE Program's Implementation Office. The GLOBE Program offers more than 50 data collection protocols with the clouds protocol being historically the most popular protocol. In 2016, The GLOBE Program launched the GLOBE Observer (GO) app. This change expanded the program from being used by teachers and students to being open to the public. The GO app includes four protocols: clouds, mosquito habitats, land cover, and tree height (Amos et al., 2020). In 2017, theNASA GLOBE Clouds team at NASA Langley Research Center began matchings each possible cloud observation submitted to GLOBE to multiple satellite data, which continues through the present day at the time of this writing.

2.1 GLOBE citizen scientist reports

GLOBE data are collected from two main sources: traditional pen-and-paper forms and through the GO mobile app (GLOBE, 2021b). The GO mobile app provides users with a convenient guided reporting tool that can be used with most modern mobile devices. First the participant provides a report of cloud and sky conditions that they observe. Second, the participant can optionally take photographs of the sky in each cardinal direction, upward, and downward, guided by the app. Finally, the participant is then able to send in their report to the

GLOBE Program.The GLOBE datasets contain a number of variables reported by citizen scientists that characterize various aspects of cloud, sky, and surface conditions. The variables of interest in this project are the cloud cover and opacity values by atmospheric level. Cloud cover is reported using discrete categories that correspond with estimated percentages of coverage. These categories are NON (0% cover), FEW (less than 10%), ISO (10% to 25%), SCT (25% to 50%), BKN (50% to 90%), and OVC (more than 90%). Cloud opacity is reported in three categories: transparent, translucent, and opaque. These estimates are qualitative in nature and can depend on a number of factors such as the darkness of the cloud base and whether the disk of the sun is visible through the clouds. NASA GLOBE Clouds aims to make the categories consistent with satellite thresholds; opaque clouds have optical depths greater than 10, transparent clouds are less than three, and translucent clouds are in-between. Chambers et al. (2017) examined cloud opacity from S'COOL data, the predecessor of NASA GLOBE Clouds, and found that compared with satellite data, transparent reports correspond with an average cloud optical depth of 8.26, translucent of 13.89, and opaque of 26.96. However, note that these results do not account for cloud type or level, so these numbers should be interpreted as general estimates rather than exact quantities.

We focus on cloud cover and opacity reports by atmospheric level. NASA GLOBE Clouds uses approximate thresholds for high, middle, and low clouds derived from the World Meteorological Organization's (WMO's) definitions: high clouds are above 6 km, low clouds are below 2 km, and middle clouds are in-between. Obviously, cloud altitude is not directly measured by GLOBE participants, but rather must be categorized into high, middle, and low clouds from what the participants see, specifically the cloud morphology. GLOBE provides participants with a number of tutorials and guides to help the participants estimate cloud level. One of the most important and useful methods is identifying cloud types. GLOBE uses the ten basic cloud types commonly used by the WMO. Cirrus, cirrostratus, and cirrocumulus are high clouds; altocumulus, altostratus, and nimbostratus are middle clouds; cumulus, stratus, and stratocumulus are low clouds. Cumulonimbus clouds are a special case because while satellites classify them as high clouds, from the ground they are considered low clouds. Note that even in cases when participants do not specifically report cloud type, they still often use cloud type as a guide to identify level.

Cloud cover reports can be difficult for participants to make. Identifying cloud level can be straightforward when only one cloud type dominates, such as cumulus or cirrus, or when two distinctive cloud types coexist, such as cumulus clouds under a cirrus layer. Pictorial examples are available within the GO app to select the best category. More complex skies will understandably result in some ambiguity in cloud height. Multiple resources, including dichotomous keys and cloud charts, have been developed to help citizen scientists identify cloud types (GLOBE, 2021c). In 2020, GO released the Clouds Wizard that guides app users through a series of questions based on the appearance of clouds to the selection of the correct cloud type (GLOBE, 2021d). The Clouds Wizard includes cloud cover and cloud opacity interactives for participants in need of assistance. For other groups like teachers and professionals, live training sessions are available as well as electronic slides that go over each portion of the protocol. Because of these possible uncertainties, in our analysis, we will examine the effect that overlapping cloud layers may have on the results.

2.2 The 2018 Spring Cloud Challenge and 2019 Fall Cloud Challenge

NASA GLOBE Clouds has been continually collecting citizen science reports and matching them with satellite data since 2016. In addition, during this time period they have conducted three intensive observation periods (IOPs): SCC18 occurring from 15 March to 15 April 2018, FCC19 from 15 October to 15 November 2019, and the 2020 Community Cloud Challenge (CCC20) from 15 July to 15 August 2020. Each IOP resulted in a significant boost in the monthly number of reports above the baseline reporting rate of about 15,000 per month. Table 1 shows the number of reports from SCC18 and FCC19, including specifically reports of cloud cover by level. Figure 3 shows the locations of reports contributing to SCC18 and FCC19. All three IOPs were worldwide events, with public participation being advertised and encouraged in many GLOBE-participating nations. Thus the IOPs represent a large sampling of many regions of Earth, allowing for regional as well as global analysis of cloud properties.

In this paper we examine data from SCC18 and FCC19. Data from CCC20 are currently being assessed at the time of this writing. A detailed description of SCC18 is given by Colón Robles et al. (2020). FCC19 was conducted in a similar manner to SCC18. Although FCC19 did not have as much media coverage, it resulted in over 45,300 cloud observations in more than 17,000 locations worldwide (Atkinson, 2019). Reports included haboobs and other types of dust events, as well as bushfires in Australia that were occurring in November 2019 (Voiland, 2020; Peterson et al., 2021). We use version 2.0 of the SCC18 and FCC19 datasets (Rogerson et al., 2018; Rogerson et al., 2019). Version 2.0 data includes additional satellite matches compared to version 1.0 after the window before or after a satellite overpass to match with ground data was corrected from 0.15 hours (9 minutes) to 0.25 hours (15 minutes). The changes in versions have resulted in small but mostly insignificant changes in some of the numbers characterizing the SCC18 dataset given by Colón Robles et al. (2020).

As Fig. 3 indicates, the data have been subdivided into six geopolitical regions as defined by GLOBE: Africa (specifically sub-Saharan, AFR), Asia and Pacific (ASP), Europe and Eurasia (EEA), Latin America and Caribbean (LAM), Near East and North Africa (NAF), and North America (specifically the United States and Canada, NAM). This grouping of GLOBE data is used because it provides large volumes of data from which to derive robust statistical results. In addition, observations taken from ships in the Southern Ocean near the Drake Passage are also included as a seventh region (DRK). We assess the cloud properties and satellite comparisons from these individual regions as well as the global mean. Granted, some of these regions are quite large and include multiple climate zones, such as ASP, which extends from Mongolia to New Zealand. Future research may subdivide the data instead to more closely match regional climatic variability.

288 **Table 1. Number of GLOBE participant reports by region and atmospheric level. For each**
289 **category, the top number is for SCC18 + FCC19, the middle is for SCC18, and the bottom**
290 **is for FCC19. Note that the number of high clouds and middle cloud obs. are actually the**
291 **same.**

region	all obs.	high cloud obs.	middle cloud obs.	low cloud obs
global	100623 55234 45389	79390 43291 36099	79390 43291 36099	92398 50708 41690
AFR	1179 446 733	1026 374 652	1026 374 652	1126 405 721
ASP	12171 6910 5261	10008 5852 4156	10008 5852 4156	11340 6487 4853
EEA	44110 28654 15456	33077 21555 11522	33077 21555 11522	40342 26306 14036
LAM	5668 2057 3611	4257 1678 2579	4257 1678 2579	4903 1866 3037
NAF	11882 5196 6686	11473 4910 6563	11473 4910 6563	11571 4946 6625
NAM	25480 11906 13574	19443 8869 10574	19443 8869 10574	22993 10633 12360
DRK	39 28 11	24 17 7	24 17 7	35 28 7

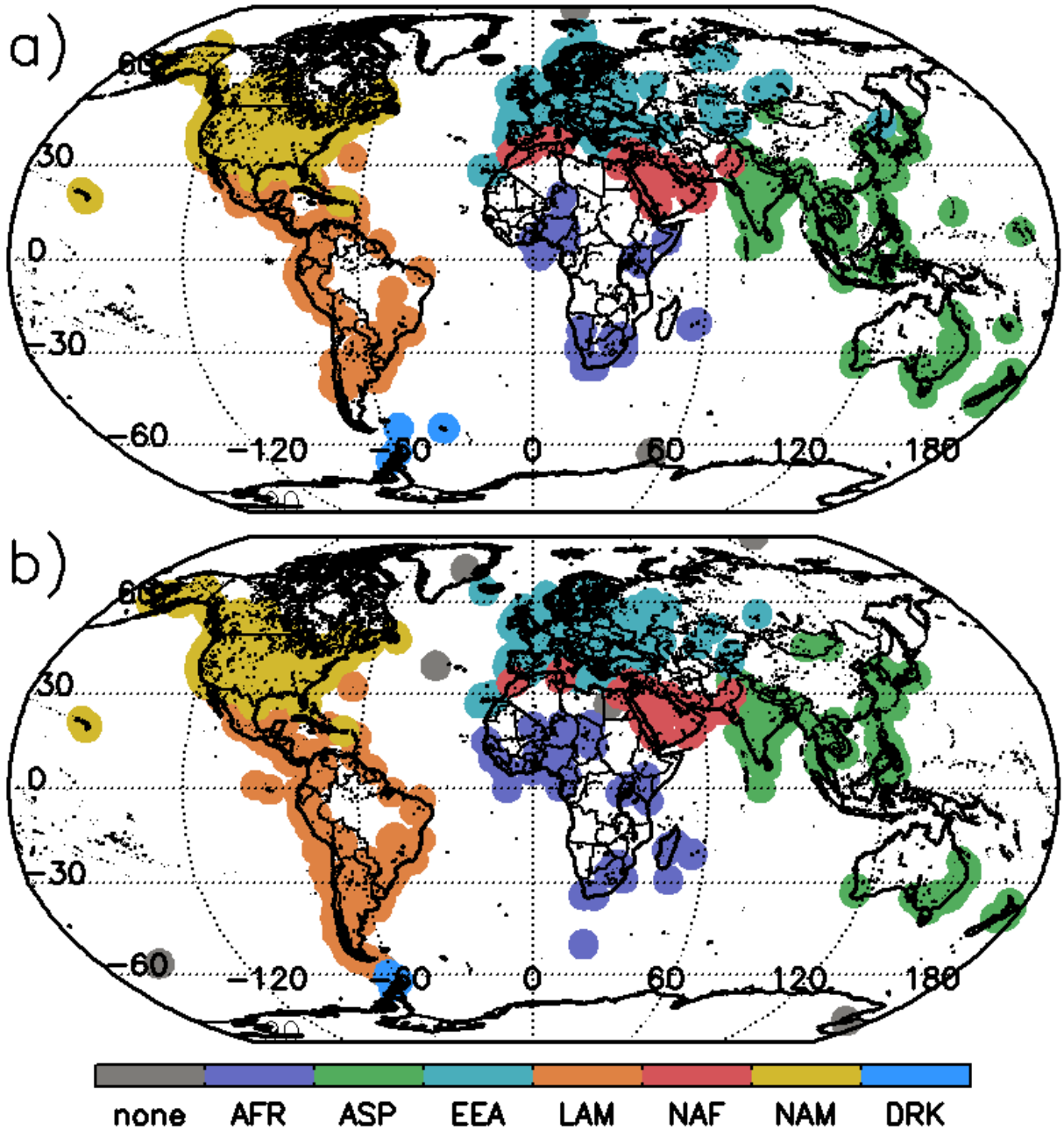

Figure 3. Global maps of observation locations for the SCC18 (top) and FCC19 (bottom). Individual points are color-coded by GLOBE geopolitical region. The regions are sub-Saharan Africa (AFR), Asia and Pacific (ASP), Europe and Eurasia (EEA), Latin America and Caribbean (LAM), Near East and North Africa (NAF), non-Latin North America (NAM), and the Drake Passage (DRK). "None" refers to data points outside the geopolitical regions.

Table 2. Number of satellite matches to GLOBE reports by region. For each category, the top number is for SCC18 + FCC19, the middle is for SCC18, and the bottom for FCC19. The number in parentheses is the ratio of satellite matches to the total number of GLOBE reports (from Table 1).

region	all satellites	GEO	Terra	Aqua	CALIPSO
global	54168 (0.54) 29997 (0.54) 24171 (0.53)	49054 (0.49) 27273 (0.49) 21781 (0.48)	6356 (0.06) 3698 (0.07) 2658 (0.06)	5542 (0.06) 3007 (0.05) 2535 (0.06)	35 (0.00) 35 (0.00) 0 (0.00)
AFR	429 (0.36) 192 (0.43) 237 (0.32)	352 (0.30) 158 (0.35) 194 (0.26)	65 (0.06) 30 (0.07) 35 (0.05)	53 (0.04) 29 (0.07) 24 (0.03)	0 (0.00) 0 (0.00) 0 (0.00)
ASP	5632 (0.46) 3530 (0.51) 2102 (0.40)	4990 (0.41) 3218 (0.47) 1772 (0.34)	504 (0.04) 273 (0.04) 231 (0.04)	568 (0.05) 297 (0.04) 271 (0.05)	2 (0.00) 2 (0.00) 0 (0.00)
EEA	17670 (0.40) 11731 (0.41) 5939 (0.38)	14561 (0.33) 9772 (0.34) 4789 (0.31)	2755 (0.06) 1855 (0.06) 900 (0.06)	2215 (0.05) 1405 (0.05) 810 (0.05)	16 (0.00) 16 (0.00) 0 (0.00)
LAM	2940 (0.52) 1285 (0.62) 1655 (0.46)	2631 (0.46) 1163 (0.57) 1468 (0.41)	277 (0.05) 133 (0.06) 144 (0.04)	269 (0.05) 123 (0.06) 146 (0.04)	3 (0.00) 3 (0.00) 0 (0.00)
NAF	5015 (0.42) 2058 (0.40) 2957 (0.44)	4746 (0.40) 1979 (0.38) 2767 (0.41)	206 (0.02) 85 (0.02) 121 (0.02)	274 (0.02) 78 (0.02) 196 (0.03)	2 (0.00) 2 (0.00) 0 (0.00)
NAM	22425 (0.88) 11174 (0.94) 11251 (0.83)	21740 (0.85) 10969 (0.92) 10771 (0.79)	2536 (0.10) 1315 (0.11) 1221 (0.09)	2147 (0.08) 1067 (0.09) 1080 (0.08)	12 (0.00) 12 (0.00) 0 (0.00)
DRK	14 (0.36) 12 (0.43) 2 (0.18)	0 (0.00) 0 (0.00) 0 (0.00)	8 (0.21) 6 (0.21) 2 (0.18)	6 (0.15) 6 (0.21) 0 (0.00)	0 (0.00) 0 (0.00) 0 (0.00)

2.3 Satellite matching

One of the main activities of NASA GLOBE Clouds is matching satellite cloud retrieval data with GLOBE participant reports. This process is described in detail by Colón Robles, Rogerson, and Dodson (2020). As Table 2 shows, 54% of GLOBE participant reports are matched with at least one satellite observation. GLOBE reports are most frequently matched with geostationary earth orbiting (GEO) satellites, as these continuously observe the same regions of Earth. The specific matched GEO satellite depends on longitude, and available satellites are GOES-15, GOES-16, Meteosat-8, Meteosat-11, and Himawari-8. A match occurs when the GLOBE report is made within 15 min of the closest observing time from the available GEO

satellite. The GOES satellites observe every half hour, and the others every hour. If there is a match in time, then the satellite footprints within 40 km of the participant are used to calculate the satellite cloud properties, provided there is sufficient usable data.

In addition, cloud data from three low Earth orbit (LEO) satellites are also matched with participant reports: Aqua, Terra, and the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO). We use Terra and Aqua in this report, as CALIPSO matches are much rarer and more difficult from which to gather robust statistics. Terra and Aqua carry the Clouds and the Earth's Radiant Energy System (CERES) instrument, which measures broadband shortwave and longwave radiative fluxes. Cloud properties are retrieved using the CERES Fast Longwave and SHortwave Flux (FLASHFlux) version 4A (NASA, 2021b). The FLASHFlux dataset is a product of the CERES project designed to process and release top-of-atmosphere (TOA) and surface radiative fluxes within one week of CERES instrument measurement. Because LEO satellites do not observe on spot continuously, the satellite matching is conducted only when a satellite footprint overlaps a participant's location within 15 min of the participant report. Because of this time requirement, participants are encouraged by GLOBE to observe near the time of LEO satellite overpasses.

2.4 Correcting the matched GLOBE data

Before vertical cloud properties can be assessed from the matched GLOBE datasets, some corrections must first be applied to the cloud cover by level data for improved consistency and robustness of the results. There are three key issues of concern here. First, the datasets do not distinguish situations where the GLOBE participants report no clouds present at a given level (i.e. "NON"), and when the participant simply declines to report the cloud cover at that level. Both situations are indicated in the datasets as missing values. This issue arises from the manner in which the GO app is constructed. Second, there are multiple scenarios when a participant may report OVC at lower levels, and some cloud cover value at higher levels. If the skies are overcast at lower levels, the participant should not be able to see clouds at higher levels, and accurately report them. Third, there are situations in which the cloud cover by level values are inconsistent with total sky cloud cover reports, e.g. when total sky cover is reported as obscured. Because of these issues, we have applied a multi-step correction to the data before conducting any analysis.

1. When the total sky cover is reported as "obscured" (i.e. hidden by fog/smoke/heavy precipitation/etc.), set cloud cover at all levels to missing values.

2. When the total sky cover is NON, and cloud cover by level are missing values, set cloud cover at all levels to NON.

3. When cloud cover at a lower level is OVC, set cloud cover at higher levels to missing values.

4. If a cloud cover at one level is reported as FEW/ISO/SCT/BKN and cloud cover at other levels are missing values, set the cloud cover at other levels to NON.

5. If cloud cover is reported at only one level, and the cover value disagrees with the total sky value, set the cloud cover at that level to the total sky value.

The matched satellite data must also be adjusted. When comparing GLOBE and satellite data, there is one additional step needed to prepare the data beyond that described in Section 2 in order to ensure as fair a comparison as possible. While the GLOBE citizen scientist cloud cover

reports are given as cloud cover categories, the satellite cloud cover retrieval values are given as percentages. To reduce any biases that may result from comparing the categorical reports from the quantitative retrievals, the satellite data are first converted to the GLOBE cloud cover categories. From here, the cloud cover categories are converted into constant percentage values: 0% for NON, 5% for FEW, 17.5% for ISO, 37.5% for SCT, 70% for BKN, and 95% for OVC. These percentages are used to calculate the mean cloud cover for both GLOBE and satellite data. The resulting satellite data are then run through a similar correction process as the GLOBE participant reports, only with the modification of assuming that higher cloud layers can block the view of lower cloud layers.

This process yields a much more consistent dataset with more robust results from the following analysis, without greatly affecting the overall results (see subsection 3.1). It also reduces any potential biases in the comparison related to the differing percentage sizes of the cloud cover categories. Of course, this correction relies on multiple subjective criteria of when to discard reported values, and future studies may improve on the correction methodology applied here.

2.5 A note about cloud cover percentages and errors

The reader should note that we will give cloud cover values as percentages, however, these should not be confused as error percentages. For example, in a hypothetical situation where the satellites report 40% cloud cover and GLOBE reports 30% cloud cover, we will characterize the disagreement as 10% (i.e. 40% - 30%), and not as a 25% disagreement (i.e. 10% divided by 40%). This is done to avoid confusion over different kinds of percentages being conflated.

3 Comparison of cloud cover between GLOBE and the satellites

3.1 Mean cloud cover globally and by region

We begin the evaluation of the GLOBE data and matched satellite observations by computing the mean cloud cover values for low, middle, and high-level clouds for both all data points and by the different geopolitical regions. Figure 4 shows the mean low, middle, and high cloud cover for GLOBE, the geostationary satellites, Terra, and Aqua. The global mean results for GLOBE (Fig. 4a) show a mean cloud cover at all levels of about 20%, with low cloud cover being about 10% greater than middle and high cloud cover. Note that these values should not be interpreted as the total sky cover, which does not account for differing cloud levels. The satellites also show a mean cloud cover by level of about 20% (again with about 10% variability by level), which is quite close to the GLOBE results.

The general agreement between GLOBE and the satellites extends to regional results as well, including the NAF region, where cloud cover is lower than the global mean because of the relatively arid climate. The general agreement is quite an interesting find, given the very different nature of the GLOBE data (collected as qualitative assessments by citizen scientist volunteers) versus satellite (quantitative retrieved measurements from calibrated instruments). This close agreement is not a mere artifact of the correction process described in subsection 2.4 (see Fig. S1). Note that other cloud variables do not agree so closely, such as cloud presence frequency (Fig. S2), where GLOBE often disagrees by 10% or more with the satellites. The close agreement about cloud cover makes further analysis of GLOBE versus satellite data easier to

conduct, as there is less need to worry about systematic biases in mean cloud cover influencing the results.

3.2 Vertical profiles of cloud cover

While GLOBE and the satellites agree overall on mean cloud cover by level, Fig. 4 suggests that there are smaller but notable differences in the mean vertical profiles of clouds as seen from the ground versus space. As discussed in Section 1, estimating the vertical profile of cloud cover can be difficult with passive satellite instruments, as low clouds can be obscured by high clouds, and optically thin high clouds may not be easily detected using either visible or infrared measurements. The data collected from SCC18 and FCC19 provide a useful ground-based information source that can be compared and contrasted with the space-based view.

Figure 5 displays the mean vertical cloud profiles from GLOBE and the satellites for global and regional domains. The results for the global domain (Fig. 5a) show that GLOBE reports about 10% more low cloud cover than the satellites, while the satellites detect about 10% (15%) more middle (high) cloud cover. In fact, the vertical profile curves for the satellites appear almost as mirror images of the GLOBE curve, flipped across the 20% cloud cover value. This mirror imaging result is possibly explained with an intuitive and straightforward hypothesis. Citizen scientists can see low clouds more easily than middle and high clouds on average, as low clouds often obscure the view of higher clouds. The opposite is true for satellites, with high and middle clouds frequently obscuring low clouds. The hypothesized relationship, and the effects of overlapping obscuring clouds, will be further investigated and quantified in Section 5.

The regional results (Fig. 5b-h) show various patterns of disagreements between GLOBE and the satellites, and they do not all resemble the mirror image pattern of the global results. Nevertheless, the satellites still generally detect less low cloud cover and more high cloud cover than reported by GLOBE participants. In fact, in regions like LAM and NAF, the differences between satellites is much greater than between GLOBE and the most agreeable satellite to GLOBE. DRK is the exception of the regions, with GLOBE showing more high cloud cover than Terra. However, the GLOBE reports from DRK comprise a relatively small number of opportunistic reports and may not represent the DRK region as a whole. Taking Fig. 5 overall into account, the low cloud/high cloud disagreement between GLOBE and the satellites appears to be a fundamental aspect of the differing perspectives from the ground versus from space. Any intercomparison study of GLOBE versus satellite cloud properties must take this difference in perspectives into account.

The satellite vertical profiles in Fig. 5a do not completely agree with each other, particularly for high clouds, where there is an 8% disagreement between Terra and Aqua. The disagreements between satellites in the various regions are even greater, with Aqua frequently reporting greater high cloud cover than Terra. While this study is not intended to be about satellite intercomparison, and there are some known slight inconsistencies between Terra and Aqua (e.g. Yost et al., 2020), perhaps some sense can be made from these results, using knowledge of the diurnal cycle of clouds. Terra (Aqua) has an equatorial crossing time of 10:30 am LST (1:30 pm LST). This three-hour difference can have a substantial effect on vertical cloud cover in convectively-active regions during boreal spring and fall. While the diurnal peak of deep convection over land occurs later in the afternoon than Aqua observes, Aqua could nevertheless capture enough convective initiation in the early afternoon, and the associated convective anvils, to influence the difference in vertical profiles between Terra and Aqua. The

disagreement between Terra and Aqua can be found for the AFR, EEA, LAM, and NAM regions, which often produce deep convection during spring and fall, but not NAF, where convection is rare. The disagreement can also be found in DRK, where convection is also rare, so this hypothesized explanation is not the only factor.

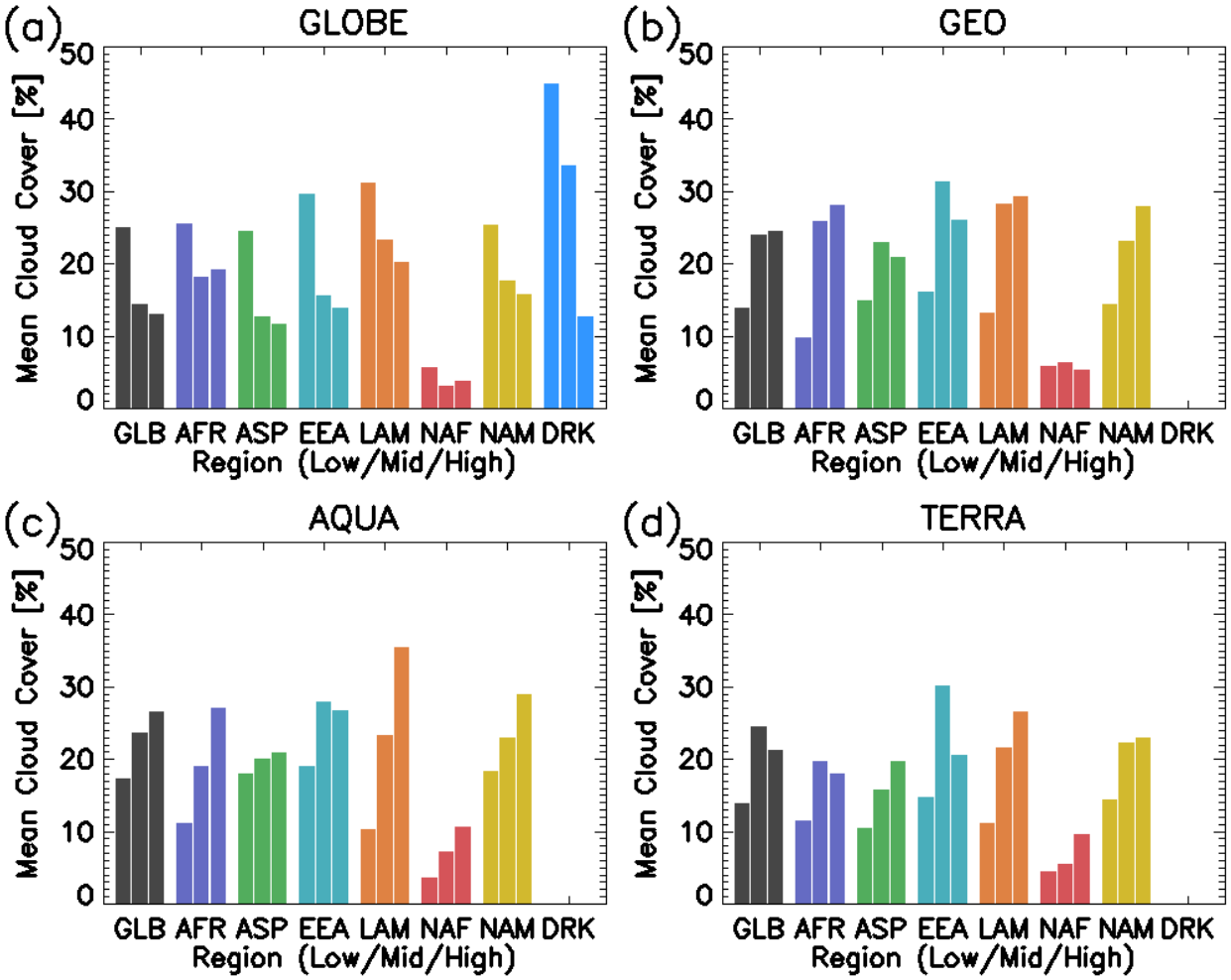

Figure 4. Mean cloud cover by atmospheric level and region for the combined SCC18 and FCC19 datasets. (a) is from GLOBE citizen scientist reports, (b) is from collocated geostationary satellite cloud retrievals, (c) is from Aqua, and (d) is from Terra. The three bars for each region represent the mean cloud cover for each level; left to right is low, middle, and high clouds.

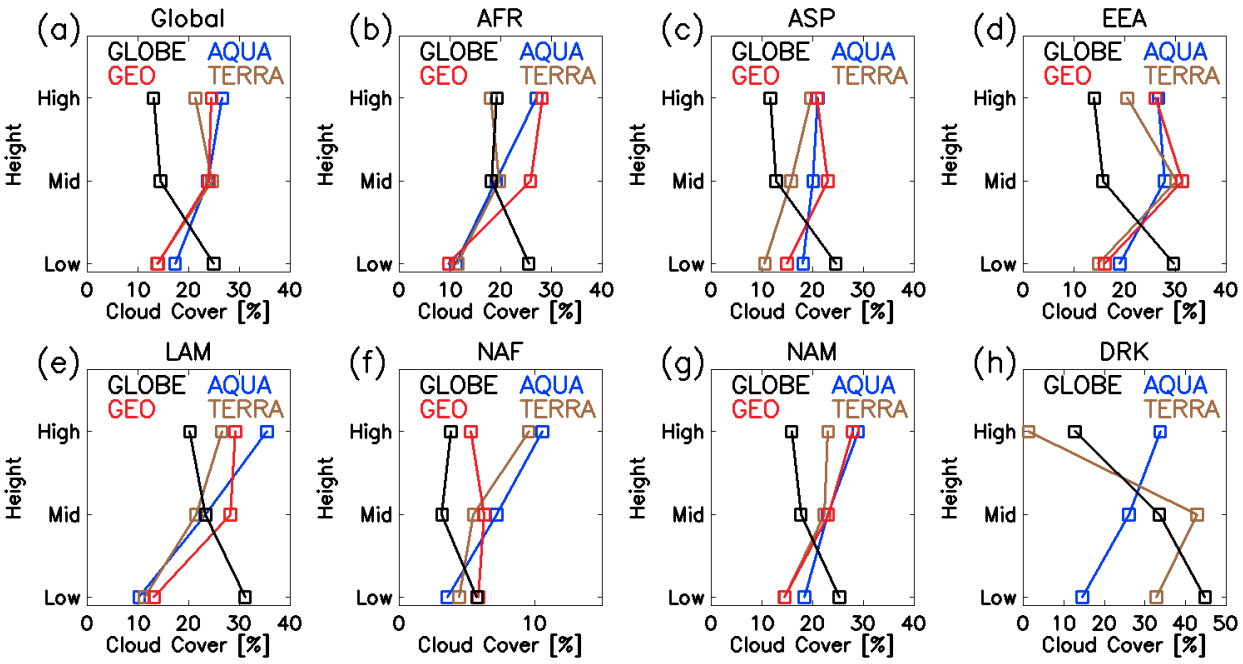

Figure 5. Vertical profiles of mean cloud cover by region for GLOBE citizen scientist reports (black), geostationary satellites (red), Aqua (blue), and Terra (brown). Note the variable range of the x axis.

4 Sensitivity of results to sample size

Given the previous comparisons between satellites and GLOBE, it might be natural to wonder how well those comparisons actually represent systematic differences between satellite observations and citizen scientist reports. After all, it is significantly more difficult to quality control subjective reports from volunteers than from purpose-built observing systems. Could these results just be a product of random noise? Can we quantitatively estimate the amount of noise in the GLOBE reports? It's reasonable to assume that the number of reports is inversely related to the noisiness of the data – increasing the number of reports should average out some of the inherent uncertainty of each report. Indeed, it is possible to quantify the effect of increased number of reports to reducing noise in the results, which is the topic of this section.

The approach to quantifying noise-induced uncertainty in the GLOBE data involves a Monte Carlo random sampling test that is adapted from Dodson et al. (2019). In this approach, a random, relatively small sample of cloud cover reports is repeatedly drawn from a large population of reports, and the mean cloud cover from the sample is compared with the population mean. By varying the sample size and running a number of sampling trials for each sample size, the relationship between the number of reports and the noise-induced uncertainty of the cloud cover calculated from those reports can be estimated. The NAM region is used because it has a large population of 25,480 reports from which to draw samples, and because using one geopolitical region reduces the noise that might be introduced from including the full range of climate zones on Earth. The specific steps of the Monte Carlo experiment are:

1. Randomly select a sample of GLOBE cloud cover reports at each level from the NAM population with a set sample size (e.g. 10 reports). The randomization does not consider specific attributes of the reports, for example whether or not a report includes cloud cover by

level data. This is to avoid biasing the statistics by favoring reports with certain attributes over others.

2. Calculate the mean cloud cover at each level from those sample reports.
3. Repeat steps 1 and 2 five thousand times for each sample size.
4. Change the sample size (e.g. from 10 to 20) and repeat steps 1-3.

Figure 6 shows the results of the Monte Carlo experiments for each level. As sample size increases, the spread of the trials decreases, as does the standard deviation of the trial values. The mean value at each sample size is almost identical to the population mean shown in Fig. 5g, consistent with the trial values being normally distributed about the population mean. This is a notable difference from the Monte Carlo results of eclipse temperature depression from Dodson et al. (2019). Also, for each sample size, low clouds tend to have a somewhat larger spread of trial values than high clouds, with a corresponding larger standard deviation (about 0.7% at 100 samples). This is true for both the SCC18 and FCC19, and suggests that GLOBE low cloud reports are somewhat noisier than high cloud reports.

To further quantify the relationship of uncertainty to sample size, Fig. 7 displays the standard deviation of the trial cloud cover results per sample size at each level, along with the mean absolute error (MAE) and the root mean square deviation (RMSD). MAE is included as an alternative to the more common RMSD, and is calculated as the average absolute value of the difference between each trial cloud cover and the population cloud cover. The use of the logarithmic axes reveal a clear relationship between sample size and uncertainty that can be described with a power law of the form $y(x) = kx^n$, where x is sample size, y(x) is the uncertainty of cloud cover, and k and n are parameters describing the shape of the curve. These parameters can be used as convenient quantitative metrics to represent the sensitivity of uncertainty to sample size. It should be noted that this power law relationship holds for RMSD and MAE only when the sample size is ~10% or less than the population mean; going much larger than 10% results in substantial deviation from a power law curve (not shown).

It is quite interesting that the exponent parameter n is almost the same, about -0.5, for all measures of uncertainty and for all levels. The parameter is slightly sensitive to whether SCC18 data is used ($n \sim -0.52$) versus FCC19 ($n \sim -0.50$). This number also holds when looking at other regions, though with more variability because of smaller population sizes. The consistency is remarkable, especially when considering the different sources of uncertainty for cloud observations for different levels (e.g. high clouds being hidden by low clouds). It suggests that the n parameter is capturing a fundamental aspect of uncertainty in GLOBE cloud reports related to sample size, rather than being sensitive to particular details of uncertainty. The k parameter, on the other hand, is sensitive to cloud level, and corresponds to the larger spread of low cloud trial values relative to high clouds.

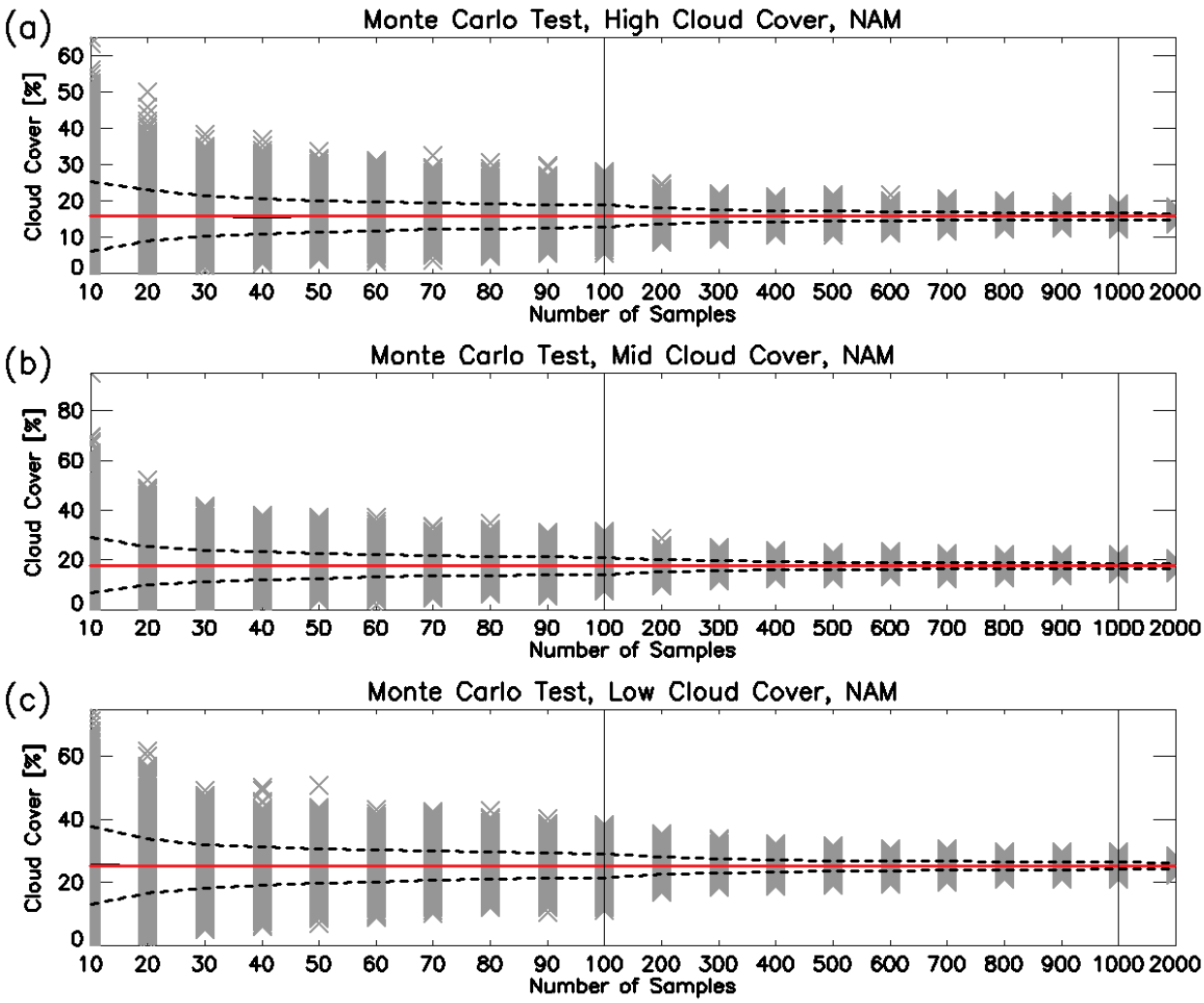

Figure 6. Sensitivity of GLOBE-reported mean cloud cover by level to sample size for the NAM region, using the Monte Carlo random sampling method. Each gray cross marks a single mean cloud cover value for one trial (out of 5000) for a given sample size (from 10 samples to 2000). The solid red line is the population mean cloud cover (from Fig. 5g), the thick black line is the mean cloud cover from all trials for each sample size, and the dashed lines are the +1 and -1 standard deviation bounds at each sample size. The vertical solid lines at 100 and 1000 indicate where the x axis changes in interval size. (a), (b), and (c) depict results for high, middle, and low clouds, respectively.

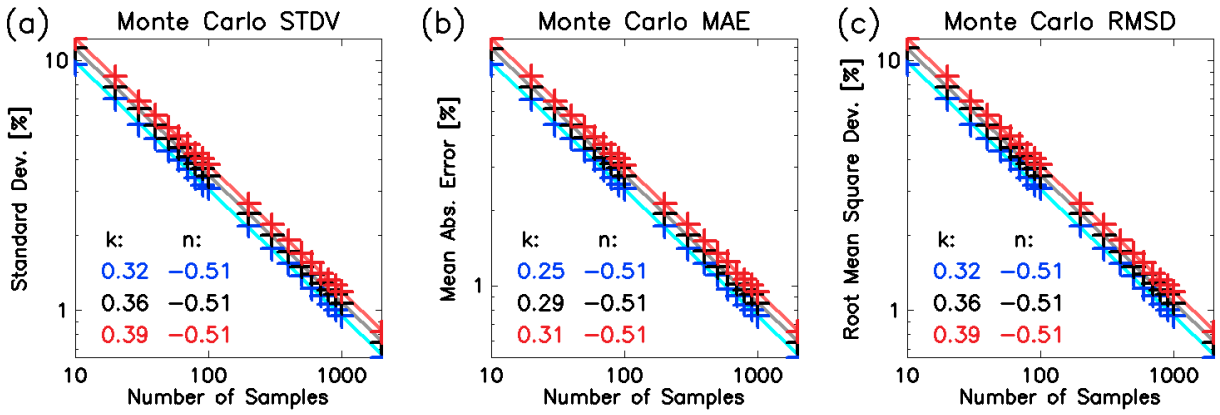

Figure 7. Sensitivity of three different measures of GLOBE-reported mean cloud cover to sample size, plotted on a log-log plot. (a) is standard deviation, (b) is mean absolute error, and (c) is root mean square deviation. Red colors are for low clouds, black/gray for middle clouds, and blues for high clouds. For each metric and level, the crosses indicate the calculated value for each sample size shown in Fig. 6, and the light colored lines show the best-fit power law curve, in which the power law function is defined as $y(x) = kx^n$. The power law values for k and n at each level are show on the lower left of each panel.

Figures 6 and 7 can be used to help understand Fig. 5. For example, Fig. 7b shows that when the number of observations is 100 (1000), the typical uncertainty in mean cloud cover associated with the number of observations is about 3% (1%). In comparison, Fig. 5 shows many disagreements between satellites and GLOBE greater than 3%. Even in GLOBE regions with smaller numbers of citizen scientist reports, such as AFR, the disagreements are often much larger than the errors associated with relatively small numbers of observations. Similarly, even the disagreements less than 10% in the NAF region cannot be entirely attributed to uncertainty from the number of observations, as the observation count exceeds 1000 for each level. So, the differences between satellites and GLOBE in Fig. 3 cannot be dismissed as contaminated by noise in the GLOBE data.

Finally, these findings show the value of increased public interest and participation in citizen science, consistent with the results of Dodson et al. (2019). Using the power law relationship, the uncertainty of mean cloud cover per level for a typical month of worldwide GLOBE data (about 15,000 reports) is about 0.29%; for SCC18 (about 55,000 reports), is 0.15%; for FCC19 (about 45,000), it is 0.17%. So the IOPs reduced the uncertainty in the results by roughly half. Increased numbers of observations clearly reduces the uncertainty of the results in a quantifiable manner to the point that scientific analysis of the data is reliable. Likewise, a scientific methodology that leverages large amounts of citizen science reports will produce more robust results than those that exclude large amounts of reports, through an overly-aggressive quality control algorithm, for example.

554 ## 5 Comparison of cloud cover by GLOBE cloud cover categories

555 ### 5.1 Cloud cover by atmospheric level

556 We now investigate more carefully how and why the GLOBE reports and satellites
557 disagree. Figure 8 displays the comparison of satellite mean cloud cover versus GLOBE reported
558 cloud cover category, separated by level. Examining Fig. 8a, the results for high clouds, two
559 aspects of the comparison are immediately apparent. First, when GLOBE participants report
560 greater cloud cover, particularly OVC, the satellites systematically under-detect cloud cover by
561 ~40% (more for Terra than Aqua and the geostationary satellites). This is a curious finding, as it
562 seems to contradict the finding from Fig. 7 that the satellites detect greater cloud cover than
563 GLOBE reports. However, it makes more sense when combined with the second aspect, that the
564 satellites over-detect clouds in lesser cloud cover conditions. This occurs particularly for NON,
565 where the satellites estimate 15% greater cloud cover than the definition value of 0%. What
566 appears to be happening is that there is a compensating effect between the overestimation of
567 satellite cloud cover relative to GLOBE in conditions with lesser cloud cover, and
568 underestimation with greater cloud cover. This compensation results in a net reduced mean high
569 cloud cover from GLOBE compared with satellites.

570 This compensating effect was not found as strongly by Dodson et al. (2019). They found
571 that satellite cloud cover when GLOBE participants reported NON was ~5%, resulting in a slight
572 net underestimate of cloud cover from satellites relative to GLOBE. The reasons for these
573 differing conclusions are not clear. The results shown in Fig.10a are found for both the SCC18
574 and FCC19, so there does not appear to be a seasonal component to the discrepancy. Future
575 efforts to compare GLOBE and satellite data may benefit from carefully testing for the presence
576 of such compensating effects.

577 Figure 8b shows a similar compensation pattern as Fig. 8a. Figure 8c, however, reveals a
578 different pattern. When GLOBE reports greater low cloud cover, the satellites greatly under-
579 detect low cloud cover relative to GLOBE. For OVC conditions reported by GLOBE, the
580 satellites underestimate low cloud cover by ~70%. The satellites appear insensitive to GLOBE
581 cloud cover reports for SCT and greater. This coupled with the same disagreement between
582 GLOBE and the satellites in lesser low cloud cover conditions, this results in the mean low cloud
583 cover reported by GLOBE being greater than the satellites; the compensating effect is shifted in
584 the opposite direction from high and middle clouds.

585

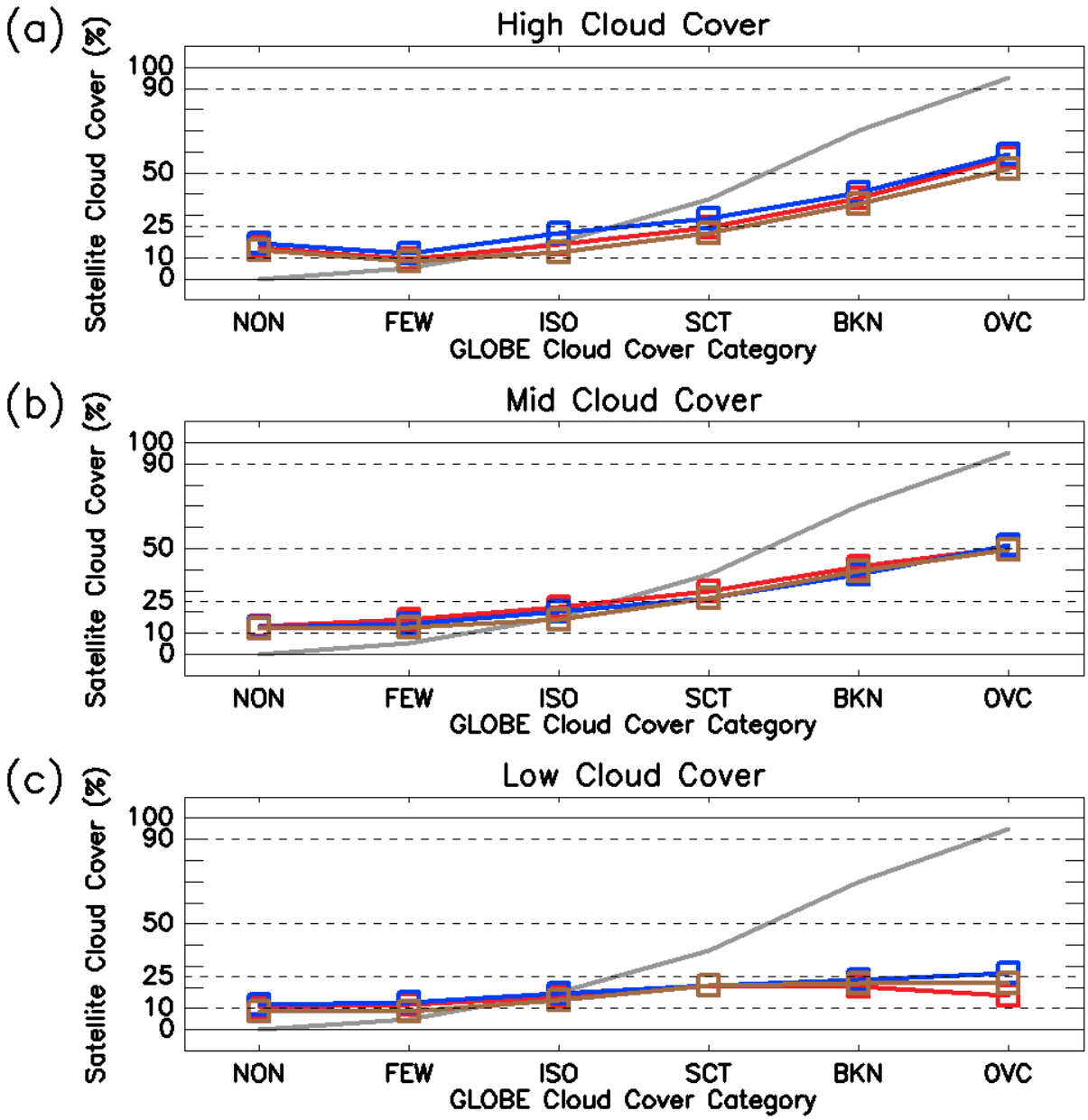

Figure 8. Mean cloud cover from satellite cloud retrievals binned by the collocated GLOBE citizen science reported cloud cover. The color curves correspond with geostationary (red), Aqua (blue), and Terra (brown). The gray curve indicates the "perfect" relationship if GLOBE agreed completely with the satellites. Dashed horizontal lines indicate the upper and lower bounds of each GLOBE cloud cover category; e.g., 25-50% are the bounds for "SCT", 50-90% are for "BKN".

5.2 Low Cloud Analysis

We previously hypothesized that the reason for the disagreement in low cloud cover between satellites and GLOBE is that the satellites' views of low clouds are often blocked by the presence of high and middle clouds. What if high and middle clouds are not present? Figure 9 repeats the calculation from Fig. 8c, but including only GLOBE reports where the satellite's

views of low clouds are not obstructed by higher clouds. Figure 9a includes only those data points in which (1) the GLOBE report gives the low cloud cover amount, and (2) in which the satellites detect 0% cloud cover at high and middle levels. The effect of this on the low cloud comparison is for GLOBE low cloud reports of SCT and greater. Aqua in particular is nearly ideal, one-to-one match with the GLOBE reports. Terra and the geostationary satellites do not agree as closely, but even the latter has a 65% improvement relative to GLOBE in OVC conditions.

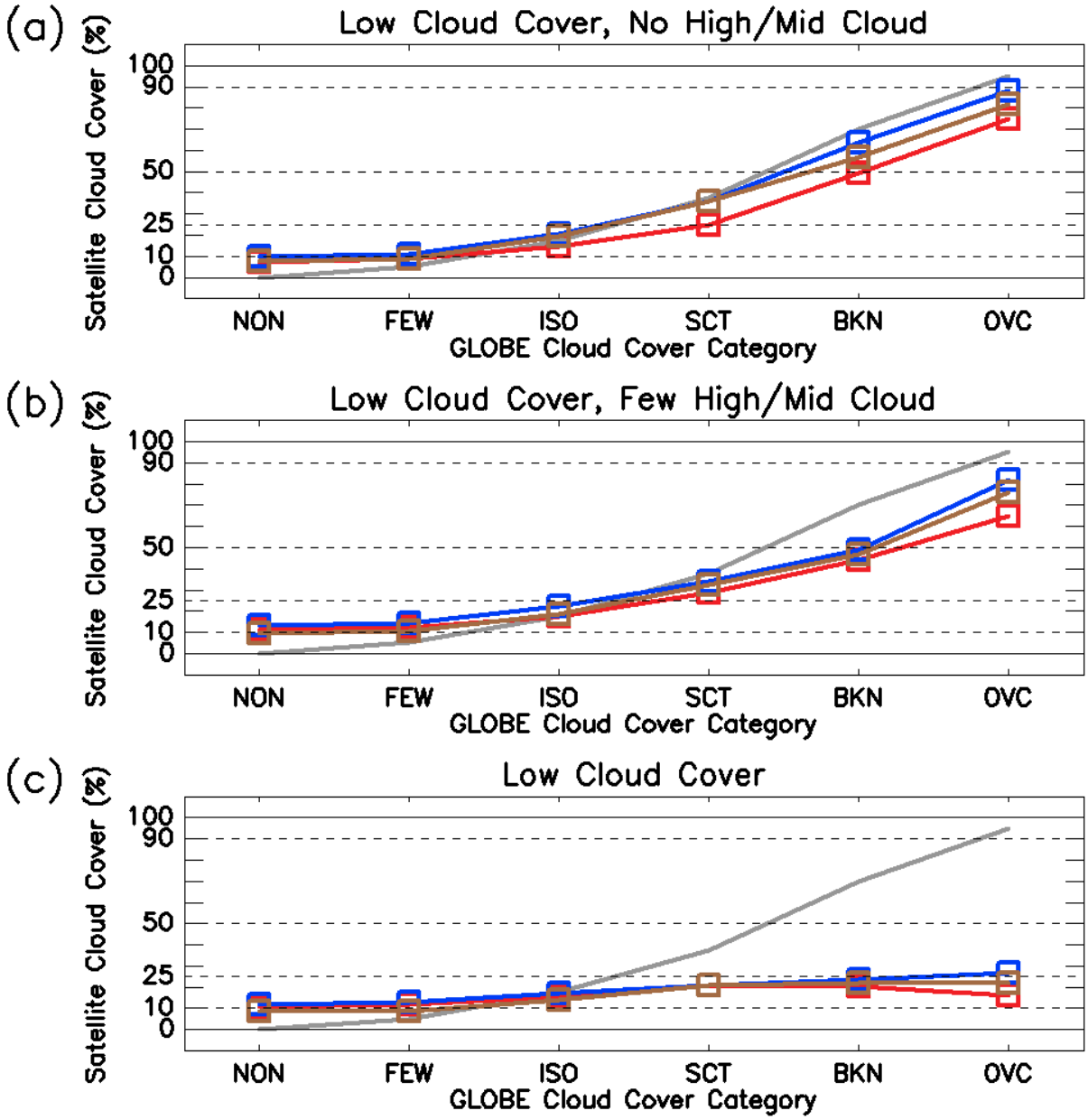

Figure 9. Mean low cloud cover from satellite cloud retrievals binned by the collocated GLOBE citizen science reported cloud cover. (a) is for data points in which the satellites detected no high- or mid-level clouds. (b) is for data points in which the satellites reported less than 25% cloud cover for high- and/or mid-level clouds. (c) is for all low cloud reports. The format is otherwise the same as Fig. 8.

Figure 9b includes only data points where the satellites report less than 25% cloud cover and high and middle levels. The agreement between GLOBE and the satellites is reduced slightly (about 10% in OVC conditions), but otherwise the agreement is still much improved over Fig. 8c. Figure 9 clearly shows that the disagreement in low cloud cover between GLOBE and the satellites is strongly primarily attributable to high and middle clouds obscuring the view of low clouds from the satellites. This result demonstrates the ability of the citizen scientist participants to accurately report low cloud cover in situations where satellite observations are unavailable.

5.3 High Cloud Analysis

The other half of the hypothesis is that the high cloud disagreement between GLOBE and satellites is that low and middle clouds partially block the citizen scientists' view of high clouds. We test this in a similar manner to low clouds, except the GLOBE reports are screened for only those with the citizen scientists reporting NON (Fig. 10a) or NON/FEW/ISO (Fig. 10b) low and middle cloud cover. Here, however, the results are quite different than for the low cloud disagreement. The curves are mostly unaffected by removing low/middle clouds, even in overcast conditions. This result holds when SCC18 and FCC19 are examined separately, as well. This is a curious and unexpected result, and merits further investigation.

Aside from overlapping clouds, there are a few other factors that could influence the disagreement in reported high cloud cover between satellites and GLOBE. Some potentially important ones include:

- The satellites may frequently miss high clouds, especially those with lesser optical depth.
- The satellites may have problems at larger viewing angles, such as at higher latitudes for geostationary satellites.
- There may be some dependence of GLOBE report uncertainty on trained versus untrained participants.

After investigating these three possibilities, the largest signal is found with cloud opacity (Fig. S3 and Fig. S4 display the results for the latter two possibilities). Figure 11 repeats the high cloud analysis, but instead with the data subdivided by GLOBE reported high cloud opacity. In OVC conditions, the satellites report 25-50% greater cloud cover for translucent versus transparent clouds (opaque high clouds reports are too infrequent to get good statistics). Terra is particularly sensitive to cloud opacity, having about twice the response of Aqua and the geostationary satellites. These results hold when examining SCC18 and FCC19 separately (not shown).

When GLOBE reports NON conditions, another major disagreement between transparent and translucent clouds occurs, where there is a large spread between the satellites for transparent clouds. In this case, Terra is in closest agreement with GLOBE, and Aqua has the largest disagreement (32%). Again, a detailed study of the disagreements between satellites is beyond the scope of this paper, but these differences may be contributing factors to the inter-satellite spread seen in Fig. 7a.

The results of Fig. 11 can inform us about the use of GLOBE reports to assess satellite observations of high clouds. In conditions with greater cloud cover, GLOBE reports may be useful for identifying situations in which the satellites miss optically thin high clouds. The

statistical approach presented here suggests there are a number of case studies that could be used for a more detailed investigation. Similarly, in lesser cloud cover conditions, GLOBE data may be used to better understand why the satellites appear to disagree so greatly. This includes the GLOBE photographic data record, while not used in this study, may be employed in an analysis that applies a case study approach.

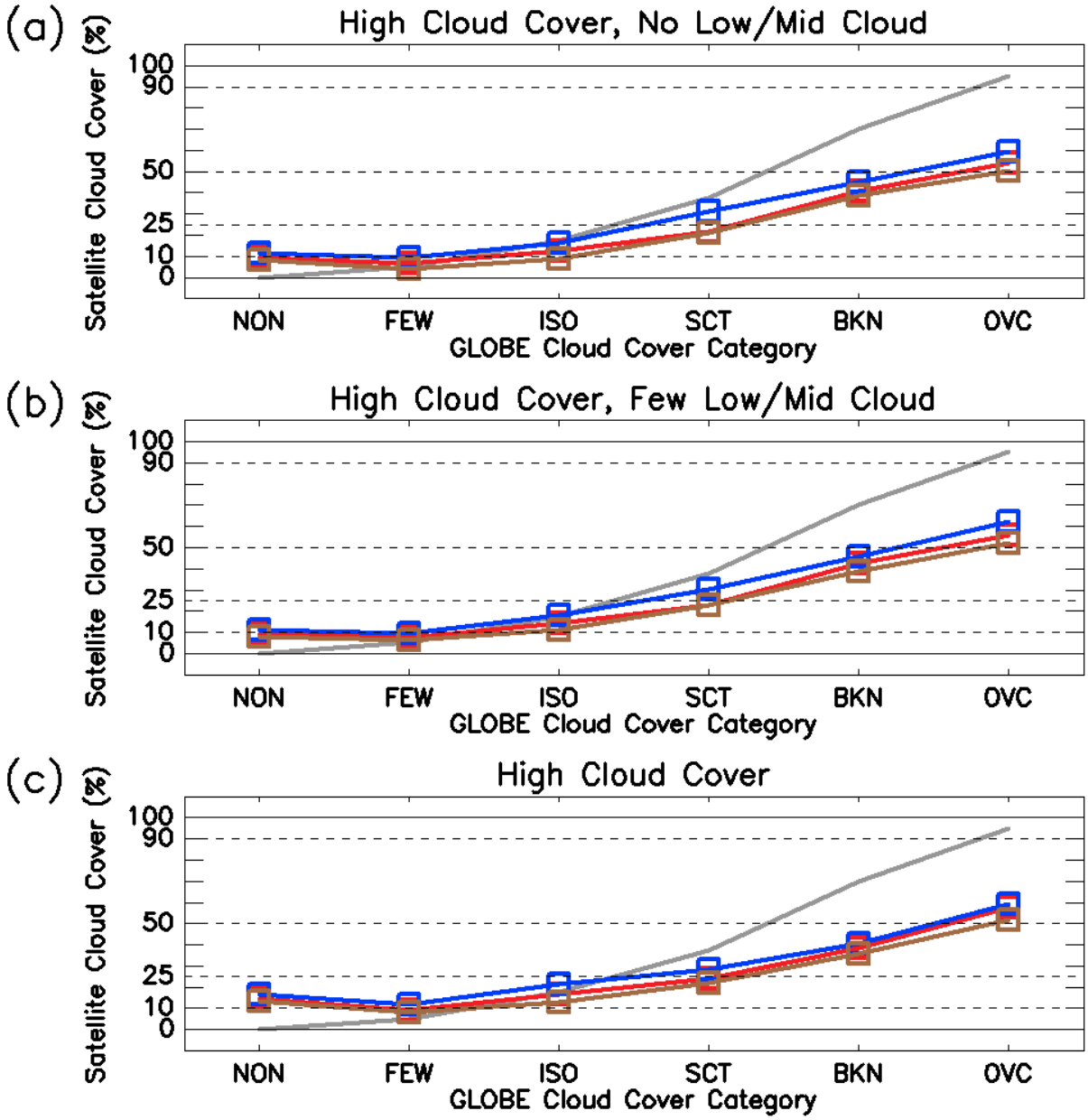

Figure 10. Mean high cloud cover from satellite cloud retrievals binned by the collocated GLOBE citizen science reported cloud cover. (a) is for data points in which the GLOBE citizen scientists reported no low- and mid-level clouds. (b) is for data points in which the citizen scientists reported "ISO" or less for low- and/or mid-level clouds. (c) is for all high cloud reports. The format is otherwise the same as Fig. 8.

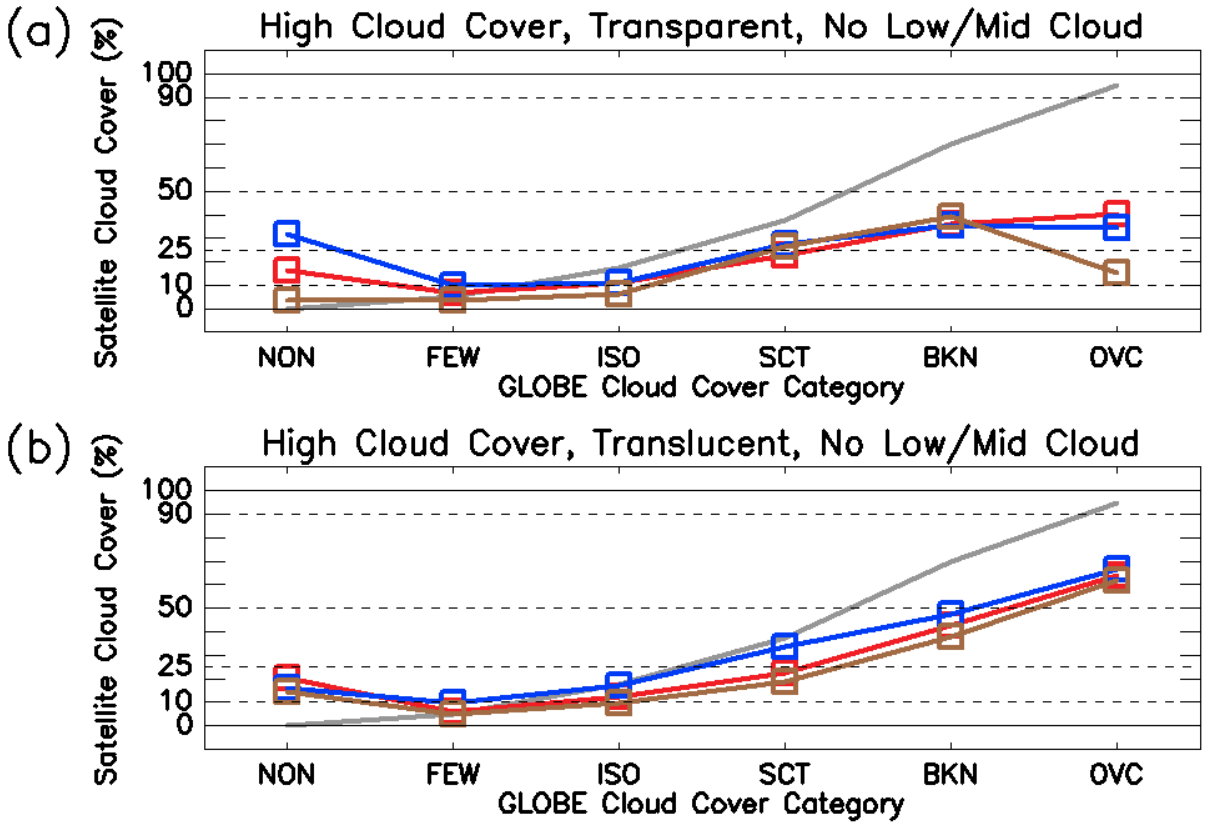

Figure 11. Mean high cloud cover from satellite cloud retrievals binned by the collocated GLOBE citizen science reported cloud cover, restricted to data points in which the GLOBE citizen scientists reported no low- and mid-level clouds. (a) is for data points in which the GLOBE citizen scientists reported transparent high-level clouds. (b) is for data points in which the GLOBE citizen scientists reported translucent high-level clouds. The format is otherwise the same as Fig. 8.

6 Conclusions

Previous studies using GLOBE data, particularly when examining the 2017 total solar eclipse, demonstrated the value of the large volumes of data collected during IOPs for producing robust scientific results from citizen science data, particularly when employing analysis methods that leverage the full volume to reduce the inherent noise of these data. Recognizing this, we employ data collected during the SCC18 and FCC19 IOPs to investigate global and regional vertical cloud structure, which has often been difficult to quantify with satellite observations. An additional benefit of using data from SCC18 and FCC19 is that with the large volume of observations, we are able to examine the robustness of the results and its dependency on the number of GLOBE participant results. This assessment gives us confidence when we further investigate the comparison of GLOBE and satellite data by stratifying the datasets by GLOBE-reported cloud cover and opacity.

The results that we have presented yield some new insights into uncertainties in the vertical profile of cloud cover estimated from satellite data, the nature of the GLOBE data, and

how citizen science may be further used in future research projects to better understand issues with satellite estimates of vertical cloud cover profiles. Specifically, they allow us to answer the three main questions asked in Section 1:

How closely does the vertical cloud structure estimated from satellite data compare with that from GLOBE data? Both global and regional mean cloud cover by level agree within 10% when calculated from GLOBE reports versus from matched satellite data. The close agreement is quite notable given the distinctly different nature of citizen scientist data versus satellite measurements, where disagreements more than 10% might be expected. However, there is a systematic disagreement between GLOBE and the satellites on the vertical structure of cloud cover, where GLOBE data consistently show greater low cloud cover, and the satellite data show greater high cloud cover. The vertical cloud cover profiles of GLOBE and the satellites are almost mirror images for the global mean, though this is not quite true for all regions.

How robust are the results from the GLOBE data? We employed a Monte Carlo test to determine the effect of sample size on uncertainty in mean cloud cover. Sample sizes of 100 result in cloud cover uncertainties of 3%, and sample sizes of 1000 result in 1% uncertainties. These uncertainties are much smaller than the disagreements between GLOBE and satellites in vertical structure of mean cloud cover, and thus show that these disagreements are not merely the result of noise in the GLOBE cloud cover reports. The relationship between sample size and random uncertainty in the results can be quantified with a power law relationship, provided that the sample size is less than 10% of the size of the full dataset. Using this power law, the increased data collection during SCC18 and FCC19 reduced the uncertainty of the results by half. The parameters of the power law may be a very useful method of quantifying and comparing the uncertainties of other GLOBE data as well as data from other citizen science projects.

If the disagreements in vertical structure between GLOBE and the satellites are not merely noise in the GLOBE reports, then what are the possible reasons for those disagreements? Partitioning the GLOBE and satellite data by GLOBE-reported cloud cover, a systematic pattern appears where satellites underestimate cloud cover in conditions of greater cloud cover, and overestimate in lesser cloud cover. The disagreement between GLOBE and satellite vertical cloud cover profiles is a product of the compensating discrepancies between greater and less cloud cover conditions. For low clouds, the satellites often miss overcast conditions because their view of low clouds is blocked by middle and high clouds. Removing situations with middle and high clouds greatly improves the comparison between satellites and GLOBE for low cloud cover. In contrast, GLOBE participants are not as severely affected by low and middle cloud cover when reporting high clouds. Instead, the disagreement between GLOBE and satellite high cloud cover is primarily sensitive to the GLOBE-reported cloud opacity, where the disagreement between GLOBE and satellites in both greater and lesser cloud cover conditions is amplified for transparent high clouds.

This last result suggests the possible merits of a more detailed study of certain cloud and sky conditions, such as transparent high clouds. One option is using the satellite-matched GLOBE data, including the matched CALIPSO observations. As mentioned in the introduction, the limited number of CALIPSO matches makes statistical analysis difficult, and so using the CALIPSO matches would need to be on a case study basis. Nevertheless the lidar carried by CALIPSO is highly complementary to the passive sensors of the other matched satellites in observing certain sky conditions such as transparent high clouds. Another option is the use of the

GLOBE photograph records. The NASA GLOBE CLOUD GAZE program (NASA, 2021a) allows citizen scientists to examine the photograph record and identify the perceived cloud and sky conditions in each photo. When completed, CLOUD GAZE will provide improved quality control of the GLOBE participant reports, as well as a valuable archive of observations as a counterpart to the matched satellite data.

Beyond addressing these specific scientific questions, the results yield some insight about the means in which the remaining issues with citizen science may be addressed. The robustness of the results show that, provided that the correct analysis method is chosen, useful scientific research can be conducted using the data gathered from a citizen science IOP. These IOPs do not have to be limited to rare geophysical events, either, in order to gather sufficient quantities of data for a reliable analysis. This reliability can be quantified and evaluated with methodologies such as the Monte Carlo random sampling method, which can go a long way towards addressing the various challenges of the inherent uncertainty in citizen science data. The careful use of IOPs and evaluation techniques can provide an avenue towards better integrating citizen science with the larger scientific endeavor for future research.

Acknowledgments

This project is based upon work supported by NASA under Award NNX16AE28A. The authors want to thank The GLOBE Program, NASA Earth Communications team, the CERES and CALIPSO satellite missions, and the NASA Langley Cloud and Radiation Research team, in particular Kris Bedka, William Smith, Douglas Spangenberg, Paul Stackhouse, and Louis Nguyen for their hard work and contributions to the collection and satellite colocation of ground-based visual cloud observations. The authors are in debt to all the citizen scientists that contributed their observations during SCC18 and FCC19. The lists of top participants can be found at https://www.globe.gov/es/news-events/globe-news/-/newsdetail/globe/nasa-globe-spring-cloud-challenge-has-wrapped-up-observation-numbers-are-sky-high-thanks-globe-community- and https://www.nasa.gov/feature/langley/globe-fall-cloud-challenge-rakes-in-the-observations. Additional thanks go to John Schimmels, David Overoye, and Travis Andersen, for maintaining the data system during SCC18 and FCC19.

Open Research

Version 2 of the GLOBE merged datasets for SCC18 (Rogerson et al., 2018) and FCC19 (Rogerson et al., 2019) are provided by GLOBE Observer and are available at https://observer.globe.gov/get-data/clouds-data.

References

Amos, H. M., Starke, M. J., Rogerson, T. M., Colón Robles, M., Andersen, T., et al. (2020). GLOBE Observer Data: 2016–2019. *Earth Space Sci.*, 7:e2020EA001175. doi: 10.1029/2020EA001175.

Atkinson, J. (2019). GLOBE Fall Cloud Challenge Rakes in the Observations. NASA. Accessed 12 July 2021. https://www.nasa.gov/feature/langley/globe-fall-cloud-challenge-rakes-in-the-observations

Barnard, L., Portas, A. M., Gray, S. L., and Harrison, R. G. (2016). The National Eclipse Weather Experiment: An assessment of citizen scientist weather observations. *Philos. Trans. Roy. Soc.*, 374A, 20150220. https://doi.org/10.1098/rsta.2015.0220.

Barker, H. W., Stephens, G. L., and Fu, Q. (1999). The sensitivity of domain-averaged solar fluxes to assumptions about cloud geometry. *Quart. J. Roy. Meteor. Soc.*, 125, 2127–2152.

Berglund, K. (1999). World Wide Weather: Involving students in GLOBE's real-life scientific research. *Science and Children*, 5. https://doi.org/10.2505/4/sc99_037_03_31.

Chambers, L. H., Young, D. F., Costulis, P. K., Detweiler, P. T., Stoddard, D. B., Sepulveda, R., Watkins, J. D., and Falcone, A. (2003). The CERES S'COOL project. *Bull. Amer. Meteor. Soc.*, 84, 759–766. https://doi.org/10.1175/BAMS-84-6-759.

Chambers, L. H., McKeown, M. A., McCrea, S. A., Martin, A. M., Rogerson, T. M., and Bedka, K. M. (2017). CERES S'COOL project update: The evolution and value of a long-running

786 education project with a foundation in NASA Earth science missions. *Bull. Amer. Meteor. Soc.*,
787 98, 473–483. https://doi.org/10.1175/BAMS-D-15-00248.1.
788 Chang, F., & Li, Z. (2005). A New Method for Detection of Cirrus Overlapping Water Clouds
789 and Determination of Their Optical Properties. *J. Atmos. Sci.*, 62(11), 3993-4009.
790 Colón Robles, M., Rogerson, T., and Dodson, J. B. (2020). NASA GLOBE Clouds:
791 Documentation on How Satellite Data is Collocated to Ground Cloud Observations (v1.1).
792 Accessed 9 September 2021, https://www.globe.gov/web/s-cool/home/satelitte-comparison.
793 Colón Robles, M., Amos, H. M., Dodson, J. B., Bouwman, J., Rogerson, T., Bombosch, A.,
794 Farmer, L., Burdick, A., Taylor, J., & Chambers, L. H. (2020). Clouds around the World: How a
795 Simple Citizen Science Data Challenge Became a Worldwide Success. *Bull. Amer. Meteor. Soc.*,
796 101(7), E1201-E1213. Retrieved Jun 4, 2021, from
797 https://journals.ametsoc.org/view/journals/bams/101/7/bamsD190295.xml
798 Dodson, J. B., Colón Robles, M., Taylor, J. E., DeFontes, C. C., and Weaver, K. L. (2019).
799 Eclipse across America: Citizen science observations of the 21 August 2017 total solar eclipse. *J.*
800 *Appl. Meteor. Climatol.*, 58, 2363–2385. https://doi.org/10.1175/JAMC-D-18-0297.1.
801 Finarelli, M. G. (1998). GLOBE: A worldwide environmental science and education
802 partnership. *J. Sci. Educ. Technol.*, 7(1), 77– 84. https://doi.org/10.1023/a:1022588216843
803 GLOBE (2021a). How cool was the eclipse? The GLOBE Program. Accessed 15 July 2021.
804 https://observer.globe.gov/hidden/science-connections/eclipse2017.
805 GLOBE (2021b). Global Learning to Benefit the Environment (GLOBE) Data User Guide, 2019,
806 version 1.0, NASA Goddard Space Flight Center. https://www.globe.gov/globe-data/globe-data-
807 user-guide.

GLOBE (2021c). Making Cloud Observations: Tips and Tricks Using the GLOBE Observer App. Accessed 16 July 2021. https://www.globe.gov/web/s-cool/home/observation-and-reporting/globe-observer-tips-and-tricks

GLOBE (2021d). Introducing the Clouds Wizard on GLOBE Observer. Accessed 16 July 2021. https://observer.globe.gov/es/news-events-and-people/news/-/obsnewsdetail/19589576/new-clouds-wiza-2.

GLOBE (2022a), observation # 909827, 15 March 2022, https://scool.larc.nasa.gov/cgi-bin/NASA-GLOBESatMatch.cgi?observation_id=116-179583-24007076-201911081325.

GLOBE (2022b), observation # 909827, 15 March 2022, https://scool.larc.nasa.gov/cgi-bin/NASA-GLOBE-CALIPSOSatMatch.cgi?observation_id=116-179583-24007076-201911081325.

Hang, Y., L'Ecuyer, T.S., Henderson, D.S., Matus, A.V. and Wang, Z. (2019). Reassessing the effect of cloud type on Earth's energy balance in the age of active spaceborne observations. Part II: Atmospheric heating. *J. Climate*, 32(19), 6219-6236. https://doi.org/10.1175/JCLI-D-18-0754.1.Hanna, E. (2018). Meteorological effects of the 20 March 2015 solar eclipse over the United Kingdom. *Weather*, 73, 71–80, https://doi.org/10.1002/wea.2820.

Hanna, E., and et al. (2016). Meteorological effects of the solar eclipse of 20 March 2015: Analysis of UK Met Office automatic weather station data and comparison with automatic weather station data from the Faroes and Iceland. *Philos. Trans. Roy. Soc.*, 374A, 20150212. https://doi.org/10.1098/rsta.2015.0212.

Horton, S.L., Clark, M.R. and Kirk, P. (2021). Investigating the Surrey tornado of 21 December 2019. *Weather*. https://doi.org/10.1002/wea.4028. Kato, S., Sun-Mack, S., Miller, W. F., Rose, F. G., Chen, Y., Minnis, P., & Wielicki, B. A. (2010). Relationships among cloud occurrence

831 frequency, overlap, and effective thickness derived from CALIPSO and CloudSat merged cloud
832 vertical profiles. *J. Geophys. Res.*, 115, D00H28. https://doi.org/10.1029/2009JD012277.
833 Kohl, H. A., Nelson, P.V., Pring, J., Weaver, K. L., Wiley, D. M., Danielson, A. B., Cooper, R.
834 M., Mortimer, H., Overoye, D., Burdick, A., Taylor, S., Haley, M., Haley, S., Lange, J., and
835 Lindblad, M. E. (2021). GLOBE Observer and the GO on a Trail Data Challenge: A Citizen
836 Science Approach to Generating a Global Land Cover Land Use Reference Dataset. *Front.*
837 *Clim.*, 3:620497. doi: 10.3389/fclim.2021.620497
838 L'Ecuyer, T.S., Hang, Y., Matus, A.V. and Wang, Z. (2019). Reassessing the effect of cloud type
839 on Earth's energy balance in the age of active spaceborne observations. Part I: Top of
840 atmosphere and surface. *J. Climate*, 32(19), 6197-6217. https://doi.org/10.1175/JCLI-D-18-
841 0753.1.
842 Li, J., Huang, J., Stamnes, K., Wang, T., Lv, Q., and Jin, H. (2015). A global survey of cloud
843 overlap based on CALIPSO and CloudSat measurements. *Atmos. Chem. Phys*., 15, 519–536.
844 https://doi.org/10.5194/acp-15-519-2015.
845 Means, B. (1998). Melding authentic science, technology, and inquiry-based teaching:
846 Experiences of The GLOBE Program. *J. Sci. Educ. Technol.*, 7(1), 97– 105.
847 https://doi.org/10.1023/A:1022592317752
848 Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., &
849 Leigh, R. R. (2015). Crowdsourcing for climate and atmospheric sciences: Current status and
850 future potential. *Int. J. Climatol.*, 35, 3185– 3203. https://doi.org/10.1002/joc.4210
851 NASA (2021a). Become an Armchair Cloud Expert and Help NASA Scientists Along the Way.
852 Accessed 14 March 2021, https://www.nasa.gov/feature/langley/become-an-armchair-cloud-
853 expert-and-help-nasa-scientists-along-the-way.

NASA (2021b). Fast Longwave And SHortwave Fluxes (FLASHflux) Clouds and Radiative Swath (SSF) TERRA-FM1 data in HDF Version 4A. Accessed 16 July 2021, https://cmr.earthdata.nasa.gov/search/concepts/C1719289020-LARC_ASDC.html

Nugent, J. (2018). Cloudy with a chance of "cirrus" science: NASA Globe Observer Clouds. *Science Scope*, 42(2), 26– 27.

Oreopoulos, L., N. Cho, and D. Lee (2017). New insights about cloud vertical structure from CloudSat and CALIPSO observations. *J. Geophys. Res. Atmos.*, 122, 9280–9300. doi:10.1002/2017JD026629.

Peterson, D. A., Fromm, M. D., McRae, R .H. D. et al. (2021). Australia's Black Summer pyrocumulonimbus super outbreak reveals potential for increasingly extreme stratospheric smoke events. *npj Clim. Atmos. Sci.*, 4, 38. https://doi.org/10.1038/s41612-021-00192-9

Portas, A. M., Barnard, L, Scott, C., and Harrison, R. G. (2016). The National Eclipse Weather Experiment: use and evaluation of a citizen science tool for schools outreach. *Phil. Trans. R. Soc. A*, 374: 20150223. http://dx.doi.org/10.1098/rsta.2015.0223

Rahman, I. U., Czajkowski, K., Jiang, Y., & Weaver, K. (2019). Validation of GLOBE citizen science air temperature observations using data from the Great American Solar Eclipse. In *Celebrating the 2017 Great American Eclipse: Lessons learned from the path of totality* (Vol. 516, pp. 501– 509). Orem: Astronomical Society of the Pacific.

Rogerson, T. M., Colón Robles, M., and Taylor, J.E. (2018). 2018 Spring Cloud Challenge v2.0 (Mar 2021 file). NASA Langley Research Center. https://observer.globe.gov/get-data/clouds-data.

Rogerson, T. M., Colón Robles, M., and Taylor, J.E. (2019). 2019 Fall Cloud Challenge v2.0. NASA Langley Research Center. https://observer.globe.gov/get-data/clouds-data.

877 Sassen, K., Wang, Z., & Liu, D. (2008). Global distribution of cirrus clouds from
878 CloudSat/Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO)
879 measurements. *J. of Geophys. Res.,* 113, D00A12. https://doi.org/10.1029/2008JD009972
880 Voiland, A. (2020). Smoky Australian Skies. NASA. Accessed 12 July 2021.
881 https://earthobservatory.nasa.gov/blogs/earthmatters/2020/01/10/smoky-australian-skies/.
882 Wang, J., and Rossow, W. B. (1998). Effects of cloud vertical structure on atmospheric
883 circulation in the GISS GCM. *J. Clim.*, 11, 3010– 3029.
884 Wang, H., Zhang, H., Xie, B. et al. (2021). Evaluating the Impacts of Cloud Microphysical and
885 Overlap Parameters on Simulated Clouds in Global Climate Models. *Adv. Atmos. Sci.*.
886 https://doi.org/10.1007/s00376-021-0369-7.
887 Wielicki, B. A., Cess, R. D., King, M. D., Randall, D. A., and Harrison, E. F. (1995). Mission to
888 Planet Earth: Role of clouds and radiation in climate. *Bull. Amer. Meteor. Soc.*, 76, 2125-2153.
889 https://doi.org/10.1175/1520-0477(1995)076<2125:MTPERO>2.0.CO;2.
890 Yost, C. R., P. Minnis, S. Sun-Mack, Y. Chen and W. L. Smith (2021). CERES MODIS Cloud
891 Product Retrievals for Edition 4—Part II: Comparisons to CloudSat and CALIPSO. *IEEE Trans.*
892 *Geosci. Remote. Sens.*, 59, 5, 3695-3724. doi: 10.1109/TGRS.2020.3015155.