Data Drought in the Humid Tropics: How to Overcome the Cloud Barrier in Greenhouse Gas Remote Sensing

C. Frankenberg[1,2], Y. M. Bar-On[1], Y. Yin[3], P.O. Wennberg[1,4], D.J. Jacob[5], A.M. Michalak[6,7]

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California, USA
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
[3]Department of Environmental Studies, New York University, New York, New York, USA
[4]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California, USA
[5]School of Engineering and Applied Science, Harvard University, Cambridge, Massachusetts, USA
[6]Carnegie Institution for Science, Stanford, California, USA
[7]Department of Earth System Science, Stanford University, Stanford, California, USA

Key Points:

- Data yields of current remotely-sensed greenhouse gas (GHG) missions in the humid tropics are often below 1%.
- Shallow cumulus clouds cause most of the low data yields, esp. in the wet season.
- Spatial resolution finer than 500 m can overcome the data sparsity in the tropics

Corresponding author: Christian Frankenberg, `cfranken@caltech.edu`

Corresponding author: Yinon M. Bar-On, `ymbaron@caltech.edu`

Abstract

Diagnosing land-atmosphere fluxes of carbon-dioxide (CO_2) and methane (CH_4), is essential for evaluating carbon-climate feedbacks. Greenhouse gas satellite missions aim to fill data gaps in regions like the humid tropics, but obtain very few valid measurements due to cloud contamination. We examined data yields from the Orbiting Carbon Observatory alongside Sentinel 2 cloud statistics. We find that the main contribution to low data yields are frequent shallow cumulus clouds. In the Amazon, the success rate in obtaining valid measurements vary from 0.1% to 1.0%. By far the lowest yields occur in the wet season, consistent with Sentinel 2 cloud patterns. We find that increasing the spatial resolution of observations to $\sim$200 m would increase yields by 2-3 orders of magnitude, and allow regular measurements in the wet season. Thus, the key effective tropical greenhouse gas observations lies in regularly acquiring high-spatial resolution data, rather than more frequent low-resolution measurements.

Plain Language Summary

Our research looks at how well satellites are able to observe greenhouse gases such as carbon dioxide and methane in tropical areas, which is important for understanding climate change. We find that these satellites often cannot make good measurements in places like the Amazon rainforest due to clouds. By using space-based instruments that can peek in between clouds (requiring about 200-300 meters spatial resolution), we would get much more frequent information, even during the rainy season. Our study shows that it's better to have high-spatial resolution, detailed satellite data regularly rather than more frequent lower resolution observations that do not yield usable measurements.

1 Introduction

While in situ measurements of greenhouse gases provide the most accurate benchmark (Komhyr et al., 1985; Andrews et al., 2014), they cannot provide spatially dense global coverage. Remotely sensed observations can't match the accuracy of in-situ measurements; however they offer the potential to provide dense spatial coverage, especially in regions where in situ measurements are limited. In the tropics, space-based measurements could enable substantial knowledge gains, as the tropics are not only sparsely sampled by in situ observations but also essential to global carbon budgets.

The tropics are, however, much more cloudy, and these clouds obscure the view from space. In passive optical remote sensing of Earth's atmosphere and surface, clouds shield the lower atmosphere and affect photon path-length distributions, greatly complicating the retrieval of greenhouse gas concentrations. This issue is particularly challenging due to the stringent accuracy and precision requirements for greenhouse gas observations (Miller et al., 2007; Merrelli et al., 2015). Consequently, rigorous cloud filtering is necessary, albeit at the cost of reducing the fraction of usable observations. Understanding the trade-off between cloud filtering and data usability is vital for assessing the scientific value of space-borne missions.

To alleviate the impact of clouds, the Orbiting Carbon Observatory (Crisp et al., 2004) utilises a pushbroom technique featuring a narrow cross-track swath width of 10 km and a spatial resolution of 1.29 km cross-track and 2.25 km along-track, which is finest resolution among existing missions targeted at atmospheric greenhouse gases. While this fine resolution was chosen to provide sufficient data even in the tropics, data yield predictions were based on cloud climatologies (Rayner et al., 2002) based on AVHRR data (James & Kalluri, 1994) aggregated to coarser scales (Stowe et al., 1999), and ignored 3D effects in the vicinity of clouds (Massie et al., 2017, 2022). The impact of small clouds ranging from tens to a few hundred meters was thus not fully captured.

Here, we revisit the impact of clouds on GHG remote sensing by quantifying long-term OCO-2 data yields. These findings are compared against cloud-free probabilities computed from 4 years of Sentinel 2 cloud data at 10 m resolution. This comparison helps us explore ways to improve the disappointing data collection from tropical regions in current satellite missions in the design of the next generation of satellites focused on space-based observations of greenhouse gases.

2 Materials and Methods

To assess the impact of clouds on greenhouse gas (GHG) remote sensing, we utilize data from the OCO-2 and OCO-3 missions (Wunch et al., 2017; Taylor et al., 2020) for actual GHG measurements and Sentinel 2 for cloud observations (Tarrio et al., 2020). OCO-2 and Sentinel 2 are on sun-synchronous orbits with an overpass time around 1:30pm and 10:30am, respectively. OCO-3 is hosted on the ISS with a precessing orbit, thus overpass times vary, enabling measurements from early morning to late afternoon (see Text and Figure S2). We find that overpass times matter somewhat, as OCO-3 data yields at the Sentinel 2 overpass time are almost a factor two higher than for the OCO-2 overpass time. Thus, the time-of-day explains some of the discrepancies between Sentinel 2 statistics and OCO-2 observed data yields noted below.

We analyze OCO-2 (v11r) and OCO-3 (v10r) data to determine the number of high quality GHG measurements, applying the 'xco2_quality_flag = 0' for accuracy. We calculate total number of measurement counts using the OCO-2's L1b files which provide a total count of the number of observations downlinked from the spacecraft. For global spatially-resolved data yields, we use the ratio of high quality (passing the quality filter) to total measurements.

Cloud statistics at coarse scales, such as those provided by MODIS, are insufficient for our analysis. Even small cloud fractions within a greenhouse gas measurement's footprint can significantly impact data quality, and are often missed by cloud climatologies. One reason for this is the stark surface albedo contrast between the O_2 A-band (about 0.4-0.5 at 760 nm) and the GHG bands (as low as 0.05 at 1.6 or 2.3 μm). Thus, a cloud with an albedo of 0.5 covering only 1% of a footprint can contribute 10% of the signal to the GHG bands but only 1% in the reference oxygen band. If this cloud shields 10% of the column (about 1 km cloud height), it can cause a low bias of 1% in retrieved gas concentrations. Thus, requiring a < ppm bias might require screening of scenes with < 0.2% fractional cloud cover. This drives our stringent 0.2% cloud fraction thresholds in the tropics.

To study the impact of clouds on OCO-2, we thus have to obtain cloud statistics at a much finer resolution than OCO-2's footprint. In the tropics, frequent shallow cumulus clouds are often linked to forest surface fluxes (Heiblum et al., 2014), are spatially organized and have cloud gaps that are smaller than 1 km in scale. Sentinel 2, with its frequent revisits and 10m resolution (Drusch et al., 2012), is uniquely suited for this task. It enables us to accurately calculate the likelihood of obtaining cloud-free measurements. In our analysis, a 'cloud free' pixel is one where less than 0.2% of the footprint area is covered by clouds or cloud shadows. We employ the Cloud Score+ product (Pasquarella et al., 2023) based on Sentinel 2 data, using a threshold of 0.65 to identify cloud-free pixels. The analysis is performed using Google Earth Engine, focusing on specific latitude and longitude ranges and using square-sized convolution kernels to vary spatial resolutions. Within each 1°x1° degree area, we compute the fraction of pixels passing our cloud filter threshold at a given footprint size for each individual Sentinel-2 image, which are acquired every five days. All individual cloud-free fractions per image are then used to compute probability distributions of cloud-fractions within a certain domain and time-period.

3 Results

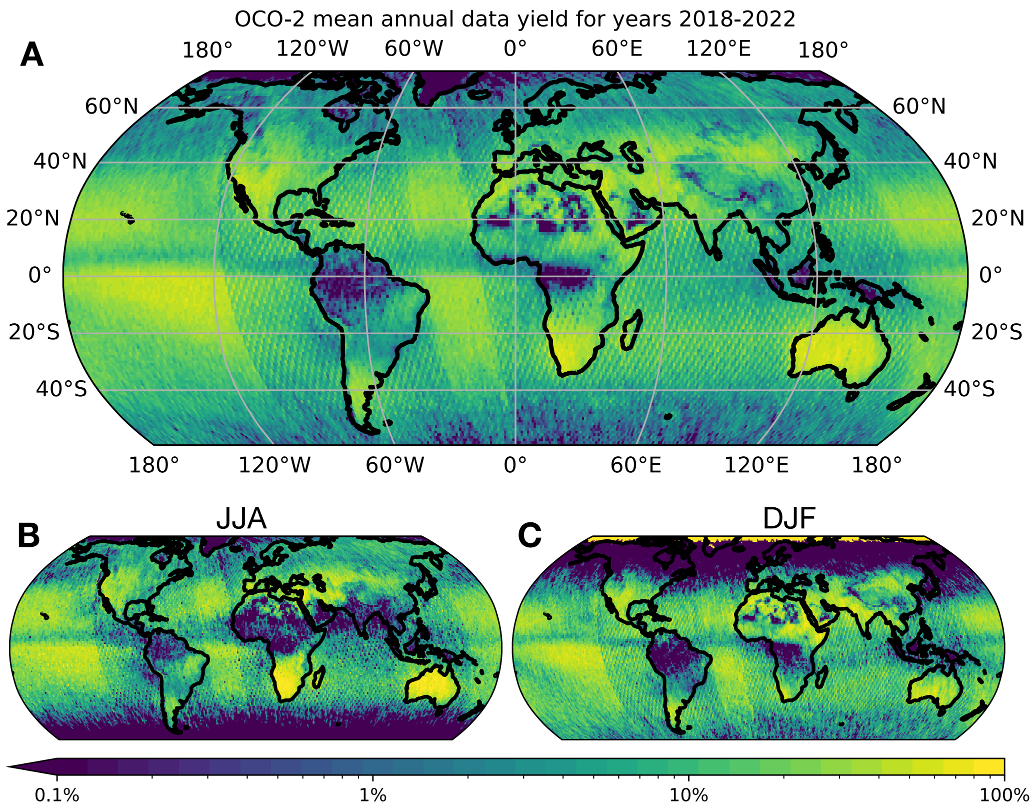

Figure 1. A) The data yield from OCO-2 within 1°x1° boxes from 2018 through 2022 on a logarithmic scale. Bottom row: Seasonal average for June/July/August (B) and December/January/February (C). The data yields vary by three orders of magnitude, with by far the lowest over areas with tropical rainforests.

Figure 1 illustrates the stark geographic variation in OCO-2 data yield based on four years of data, revealing significant disparities spanning three orders of magnitude. In humid tropical areas, data yield frequently falls below 1%, especially during the wet season. Conversely, most other global regions consistently show yields above 5%, apart from some spurious regions in the Sahara, which don't pass the OCO-2 quality flag for unrelated reasons, likely due to brightness threshold used in OCO-2 filtering. On seasonal time-scales the largest changes are due to varying solar zenith angles at high latitudes and movement of the ITCZ in the tropics.

Other peculiar patterns appear over the oceans, as orbit tracks that are predominantly over the oceans are almost exclusively performed in the glint observation mode, which has much higher data yields over dark oceans. OCO-2's orbits have a repeat cycle of 16 days, which means that a location with a 0.2% data yield would essentially never be observed (i.e., less than once every 20 years). This represents a dramatic discrepancy between potential and actual revisit times in the humid tropics. While OCO-2 can theoretically monitor tropical fluxes (Liu et al., 2017), the reliance on measurements from data-rich surrounding areas and from rare clear-sky conditions can introduce systematic biases when attempting to infer carbon fluxes with inverse analyses.

For the TROPOMI methane product (Hu et al., 2016), similar reductions in data yields have been observed, spanning more than three orders of magnitude (Qu et al., 2021). TROPOMI's larger footprint of 5-7 km allows a wide swath and daily revisit times, but might at the same time explain the even lower data yields. Thus, OCO-2 and TROPOMI data yields raise questions about the effectiveness of frequent km-scale resolution observations in data-sparse tropical regions.

The humid tropics show the worst data yields but are arguably the most important place for observing the global carbon cycle, as they have the highest above-ground carbon stocks (Santoro et al., 2020) and natural methane emissions (Saunois et al., 2020). To obtain more reliable flux estimates across major tropical areas and, more importantly, to capture spatial variations within the heterogeneous major tropical basins, higher data density both spatially and temporally is key. Towards that goal, we have to evaluate why the current missions have such low data yields and how we can mitigate this shortcoming.

To quantify the role of clouds on data yields, we derive probabilities of cloud-free satellite footprints at varying footprint resolutions using Sentinel-2 data. Due to computational demands, we focused on the tropics, the areas with the lowest data yields in OCO-2, OCO-3, TROPOMI, and GOSAT (Yokota et al., 2009).

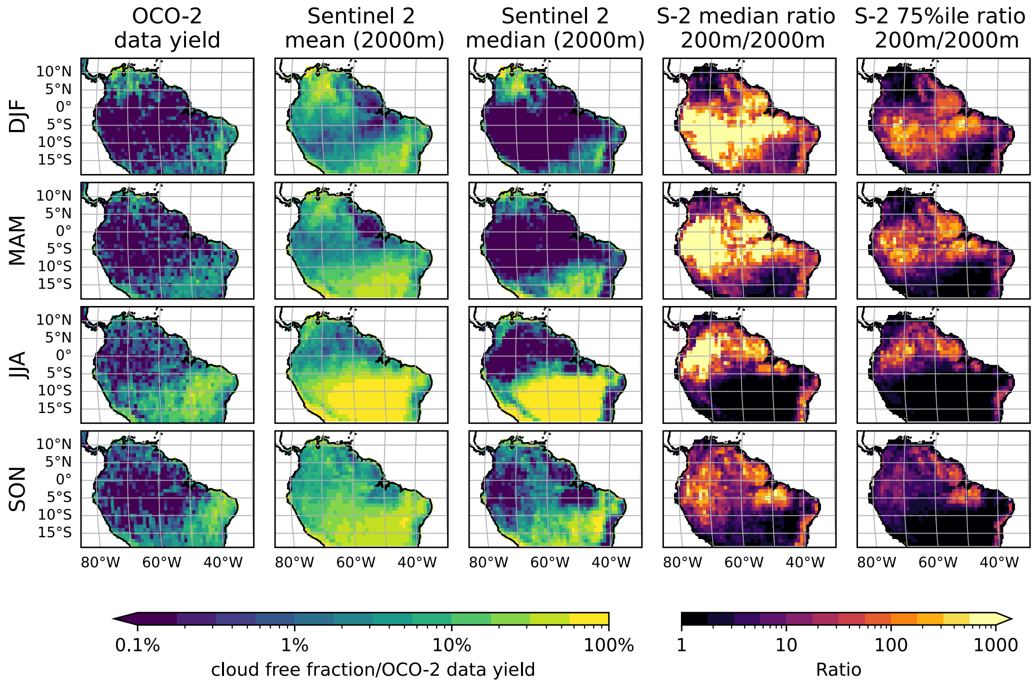

Figure 2. OCO-2 data yields and Sentinel-2 cloud statistics in South America. The rows show seasonal variations individually averaged over three months, while the columns show the mean OCO-2 data yield, the mean and median cloud free fraction at a spatial footprint of 2 km, and the ratio of the median and 75-percentile cloud-free fraction for a spatial footprint of 200 m vs. 2000 m.

Figure 2 shows the seasonal variations of OCO-2 data yields and Sentinel-2 cloud statistics in the Amazon region (see SOM for other areas). There are strong spatial and temporal commonalities between the OCO-2 data yields and the cloud statistics, confirming that clouds are a major contributor to yield reductions. A striking feature is the difference between the mean and median of cloud-free probabilities for a 2 km pixel size. This is especially true within the Amazon basin, where the mean can be more than an order of magnitude higher than the median. A 200 m resolution would increase the median of the cloud-free likelihoods by 2-3 orders of magnitude compared to 2 km pixels. The 75 percentile would increase by 1-2 orders of magnitude. Overall, it appears that OCO-2 obtains fewer valid measurements than we would expect just on the basis of clouds, missing out on the occasional cloud-free scenes that contribute dis-proportionally to the mean. It may be related to the challenge that tropical rainforests are much darker in the CO_2 bands at 1.6 and 2 μm. For instance, there are large areas below 5 degrees south in June/July/August, with relatively cloud-free conditions yet unexpectedly low OCO-2 data yields, warranting further investigations into the filter criteria employed by the OCO missions. In general, the OCO-2 yields are often more similar to the median of our cloud statistics. While time-of-day of our cloud statistics (10:30am) vs. OCO-2 (1:30pm) might be a reason for worse yields in OCO-2, it does not fully explain the discrepancies (see Text and Figure S2).

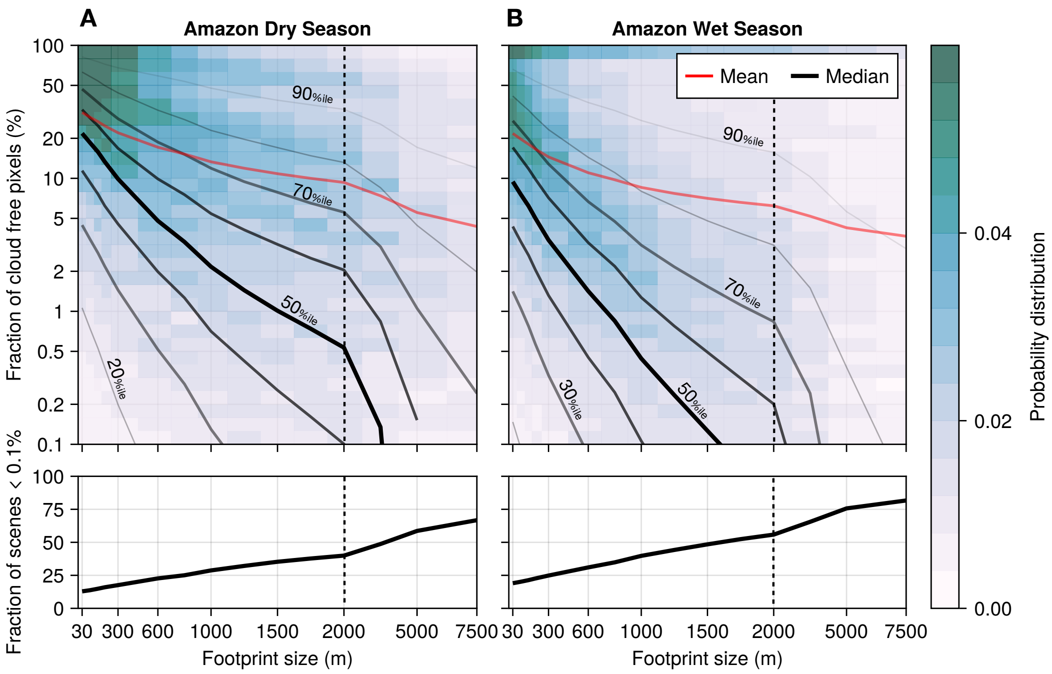

Figure 3. Frequency distribution of the fraction of cloud-free scenes for the dry (left, July through September) and wet (right, December through May) season within the Amazon (70W-60W; 2S-2N)]. Note the scale break at 2 km. The distribution on a log-scale as a function of footprint size is shown color-coded on the top row, with the percentiles (median as thick line, including 10th percentile steps with decreasing thickness) and the mean shown as lines. Towards larger footprints, the distribution gets more skewed, especially during the wet season, where the mean and 90th percentile lines intersect at around 5 km footprint size. The bottom row shows the fraction of scenes that had less than 0.1% cloud free pixels, reaching 80% in the wet season at a pixel size of 7.5 km.

A peculiar feature in the Amazon is the extreme skewness of cloud free likelihood distributions, while the mean and median are closer to each other outside the Amazon basin, and when yields are higher in general. The much higher mean indicates that a few large-scale cloud-free events contribute disproportionately to overall data yields. The median, however, is more indicative in how likely each satellite overpass is of exceeding a cloud-free fraction, i.e. there is a 50% chance of observing a higher cloud free fraction than the median.

In Figure 3, we show the probability distribution of obtaining cloud free pixels (on a logarithmic scale) as a function of footprint size in the dry and wet season, respectively. In the dry season, there is a probability peak at high cloud-free fractions, consistent across all footprint sizes. In the wet season, the peak at large cloud-free fractions diminishes and a peak around 5-30% likelihoods appears at smaller footprints, moving towards 0.5% at 2 km footprint size. The median drops by two orders of magnitude moving 30 m to 2 km footprints. In both seasons, the mean and median diverge with footprint size. In the wet season, the mean intersects the 90th percentile at 5 km footprint size, underlining that rare events dominate the mean with increasing pixel size. Also, there are a few scenes available with very high cloud free fractions, likely caused by large-scale subsidence. These few scenes contribute substantially to the mean data yield, which decays much less with footprint size than the median or the percentiles. In addition, some of the regions where the ratio in the median and 75%ile is not as enhanced in the middle of the Amazon in Figure 2 are associated with rivers or open water, which can cause large-scale subsidence and cloud-free scenes but which are dark in the nadir observation geometry. Thus, some of these scenes might be filtered out in OCO-2 because of the lack of reflected light from open water surfaces.

In principle, we can leverage the finding that smaller footprint sizes can dramatically increase the fraction of cloud free pixels. However, smaller pixels will be more noisy and limit the theoretical revisit time in the absence of clouds, as it is harder to feature a wide swath while having small pixels. As long as measurements are photon shot noise limited, however, the noisiness of individual pixels is less of a concern, as we can aggregate valid pixels. In this case, aggregating valid smaller pixels within a larger domain to cover an integrated surface area of 2x2 km^2 will have the same precision as one single cloud-free measurement at 2x2 km^2 footprint size within that larger domain. In the case of shallow cumulus cloud fields, many small pixels within cloud gaps can be aggregated while almost all large-footprint measurements will have failed due to fractional cloud cover. Thus, the increase in data yield with footprint size strongly depends on the spatial scales of clouds and gaps in between, which can be highly structured in tropical forests (Heiblum et al., 2014).

Thus, a hypothetical instrument that averages measurements over a given surface area will have the same precision irrespective of footprint size. If we define a threshold for the cloud free fraction that enables enough measurements within a larger geographic domain, we can derive how often and how evenly spaced in time an area will be observed depending on the footprint size. Here, we chose a 2% cloud fraction cutoff, which is equivalent to a single 2000 m footprint pixel within a 20x10 km^2 domain.

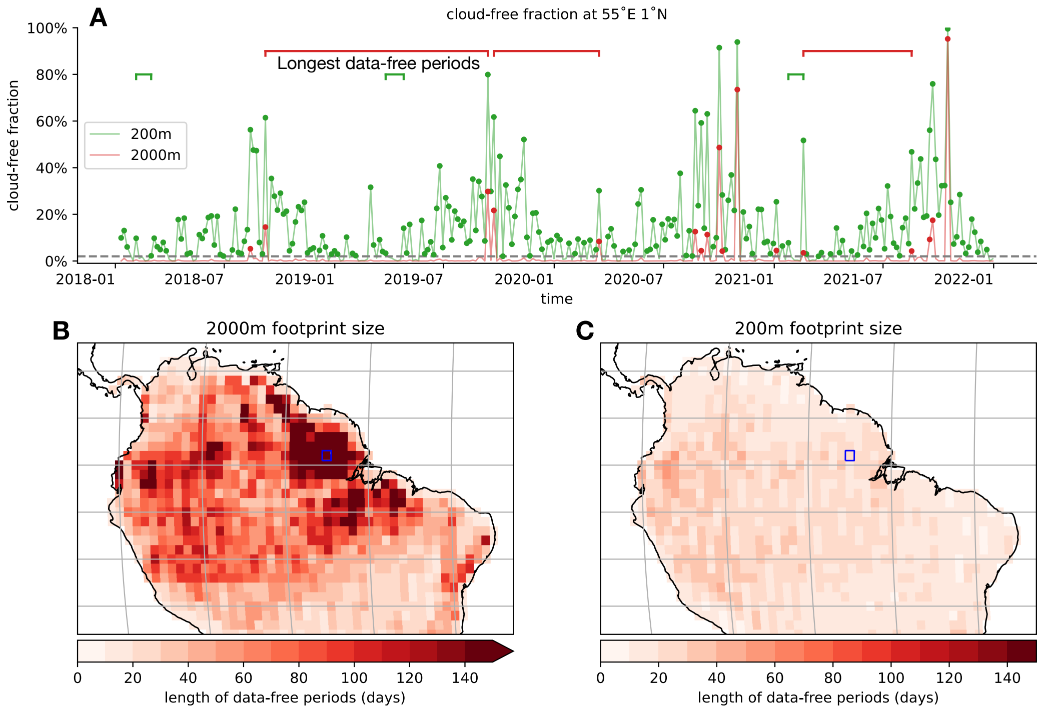

Figure 4. A) Time series of the cloud-free fraction of individual Sentinel-2 images (every 5 days) for a 200 m and 2000 m footprint resolution. Dots mark scenes that exceed the 2% threshold and the bars outline the longest three periods without any valid data acquisitions passing this threshold. B) Map of the average of the three longest data-free periods at 2000 m footprint size (in days). C) as B) but for 200 m footprints, 5-10 times shorter than for 2000 m. At smaller footprint sizes, more data is acquired over and spaced over more regular time intervals, even in the wet season.

The skewness in the cloud statistics already hinted at the fact that rare events dominate the mean at coarser spatial resolutions, while most observations have much lower cloud free scenes. This should increase the time-period in between useful measurements. Figure 4 shows a time-series of cloud free fractions computed from individual Sentinel-2 images within a 1x1 degree area in the Amazon. The very skewed distribution at 2000 m footprint size can be observed: Most scenes are below the 2% cutoff and many that pass the threshold have very high cloud free fractions, indicating the absence of broken cloud fields. At 200 m footprint size, the distribution of the cloud-free fraction is more evenly distributed and no long periods devoid of any useful measurements exist. Panels B-C show the average length (in days) of the three longest data-free periods at 2000 m and 200 m resolution, respectively. The data free periods (at 2000 m) exceed 150 days in many regions within the Amazon, even if sampled every 5 days as Sentinel 2 does. Thus, it appears unlikely that even daily revisit times in the Amazon at coarse spatial resolution would provide shorter data-free periods. Somewhat contrary to intuition, the key to observing the humid tropics more frequently is thus not to have more frequent measurements but to have finer spatial resolution (see also Fig. S3 for the trade-off between revisit times and footprint sizes). Given the importance of the tropics for GHG fluxes and the currently poor revisit times for valid measurements in these regions, prioritizing high resolution greenhouse gas measurements is needed.

4 Discussion and Conclusions

To minimize the impact of prior assumptions on estimation of GHG fluxes, inverse methods require dense measurements in both space and time. In the tropics, in situ observations are very sparse and provide motivation for using remote sensing from space to fill in the gaps. To date, however, GHG missions have had little success in observing CO_2 and CH_4 above tropical forests, with the fraction of valid retrievals varying over space and time and being as low as 0.1% during the wet season.

Using Sentinel 2 data at 10 m resolution, we illustrate that the Achilles heel of tropical remote sensing of GHGs is clouds. For example, at the footprint size of OCO-2, we find that the likelihood of obtaining a cloud-free satellite footprint is low and the distribution of likelihoods is highly skewed. Generally, both large-scale cloud systems or shallow-cumulus cloud fields reduce the likelihood of observing cloud-free scenes by several orders of magnitude, such that most satellite orbits passing the humid tropics yield almost no measurements. Periodically, however, the observations coincide with large-scale subsidence, suppressing clouds and contributing most of the valid measurements; this, in turn, might bias the flux inversion. Thus, mean data yields can hide the fact that there can be prolonged data-free periods.

The choice of equator crossing time also plays a role in limiting data. While convective systems and their associated clouds typically follow solar heating – increasing in the late morning and peaking in the early to late afternoon, our analysis with time-of-day resolved statistics from the OCO-3 mission, suggests maximum data yield just before mid-day likely because the lower solar zenith angle reduces the spatial domain through which the direct light-path traverses and lowers the amount of shading in between the clouds (Text and Figure S2).

What spatial resolution is ideal for measurements in the tropics? Even though data yields improve with smaller pixel size, the revisit time increases, thereby offsetting some of the benefit of the high spatial resolution. We find that footprint sizes around 200-400 m, can optimize coverage and revisit times, potentially solving the data-drought problem in the humid tropics (Figure and Text S3).

To improve greenhouse gas remote sensing in the tropics in the future, we need dramatically increase the amount of measurements that are not impacted by clouds. Thus, we need either more cloud-free observations or use algorithms that are less sensitive to fractional cloud cover. For cloud-avoidance, better spatial and temporal sampling requires much better spatial resolution than currently available. How can we leverage that finding? Obtaining high spatial and spectral resolution using passive spectroscopy is difficult, but recent studies have shown that very high spectral resolution is not necessarily required (Cusworth et al., 2019; Jongaramrungruang et al., 2021; Wilzewski et al., 2020; Galli et al., 2013). In terms of algorithm choices, a major complication for the operational OCO-2 retrievals is that the contrast between cloud and surface albedo within the O_2 A-band and the CO_2 bands is very different. As the radiative transfer forward model requires horizontal homogeneity, sub-pixel clouds violate that requirement and can thus be very sensitive to fractional clouds, as the relative contribution from clouds to back-scattered light will be much higher in the CO_2 bands. For complex retrieval methods that rely on a separate oxygen band to constrain the light-path distribution, this can be critical. Methods that can use a proxy gas for light path referencing within the same

—or nearby— wavelength window as the target gas of interest avoid that problem. Thus, these methods are less sensitive to clouds and can make use of simpler radiative transfer schemes. Such a proxy retrieval has been successfully implemented with SCIAMACHY, which measured at 1.35 nm FWHM (Frankenberg et al., 2005) and yields results comparable to more complex algorithms (Schepers et al., 2012). Similarly, a proxy retrieval substantially increased data yields in the tropics for the GOSAT mission (Parker et al., 2020) as cloud filters could be relaxed. However, these methods so far rely on CO_2 measurements as a proxy gas; thus, a different proxy gas would be required to constrain CO_2 itself. N_2O could be a viable alternative as it varies much less in the troposphere than either CH_4 and CO_2 and has absorption features in the vicinity of the strong CO_2 and CH_4 bands within the 2-2.4 μm range. Active systems with small footprint are also possible, especially as they observe in a true nadir geometry, both for illumination and receiver, eliminating cloud shadows and maintaining a constant viewing geometry across the globe. Lidar observations thus provide another path forward, as the laser pulses typically have footprints of less than 100 m, thus also being able to observe in between clouds or over fully cloudy pixels, as this fine spatial resolution greatly reduces horizontal heterogeneity in both surface and cloud properties (Ramanathan et al., 2015; Ehret et al., 2017; Mao et al., 2018). In essence, small ground pixels are key for solving the data drought in the humid tropics. They offer the additional advantage of observing stronger spatial concentration gradients in greenhouse gases, which improves flux inversions.

5 Open Research

5.1 Data Availability Statement

All datasets used in this manuscript are publicly available and archived either through NASA data centers or Google Earth Engine. OCO-2 and OCO-3 XCO_2 lite files are available from the DAAC archive (OCO-2/OCO-3 Science Team, Payne, & Chatterjee, 2022; OCO-2/OCO-3 Science Team, Chatterjee, & Payne, 2022). OCO-2 and OCO-3 L1b files are are available from the same DAAC (OCO-2 Science Team et al., 2022, 2022). Cloud identification data based on Sentinel 2 are obtained through the Cloud Score+ product (Pasquarella et al., 2023), publicly available as Image collection at `https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_CLOUD_SCORE_PLUS_V1_S2_HARMONIZED`.

Acknowledgments

We thank the Google team to not only provide the Google Earth Engine but also the CloudScore+ cloud product, which greatly facilitated our work.

References

Andrews, A., Kofler, J., Trudeau, M., Williams, J., Neff, D., Masarie, K., ... others (2014). CO_2, CO, and CH_4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: Instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts. *Atmospheric Measurement Techniques*, *7*(2), 647–687.

Crisp, D., Atlas, R. M., Breon, F.-M., Brown, L., Burrows, J., Ciais, P., ... others (2004). The orbiting carbon observatory (OCO) mission. *Advances in Space Research*, *34*(4), 700–709.

Cusworth, D. H., Jacob, D. J., Varon, D. J., Chan Miller, C., Liu, X., Chance, K., ... others (2019). Potential of next-generation imaging spectrometers to detect and quantify methane point sources from space. *Atmospheric Measurement Techniques*, *12*(10), 5655–5668.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... others (2012). Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, *120*, 25–36.

Ehret, G., Bousquet, P., Pierangelo, C., Alpers, M., Millet, B., Abshire, J. B., ... others (2017). Merlin: A french-german space lidar mission dedicated to atmospheric methane. *Remote Sensing*, *9*(10), 1052.

Frankenberg, C., Meirink, J., van Weele, M., Platt, U., & Wagner, T. (2005). Assessing methane emissions from global space-borne observations. *Science*, *308*(5724), 1010–1014.

Galli, A., Guerlet, S., Butz, A., Aben, I., Suto, H., Kuze, A., ... others (2013). The impact of spectral resolution on satellite retrieval accuracy of co 2 and ch 4. *Atmospheric Measurement Techniques Discussions*, *6*(6).

Heiblum, R. H., Koren, I., & Feingold, G. (2014). On the link between amazonian forest properties and shallow cumulus cloud fields. *Atmospheric Chemistry and Physics*, *14*(12), 6063–6074.

Hu, H., Hasekamp, O., Butz, A., Galli, A., Landgraf, J., Aan de Brugh, J., … Aben, I. (2016). The operational methane retrieval algorithm for tropomi. *Atmospheric Measurement Techniques*, *9*(11), 5423–5440.

James, M. E., & Kalluri, S. N. (1994). The pathfinder avhrr land data set: An improved coarse resolution data set for terrestrial monitoring. *International Journal of Remote Sensing*, *15*(17), 3347–3363.

Jongaramrungruang, S., Matheou, G., Thorpe, A. K., Zeng, Z.-C., & Frankenberg, C. (2021). Remote sensing of methane plumes: instrument tradeoff analysis for detecting and quantifying local sources at global scale. *Atmospheric Measurement Techniques*, *14*(12), 7999–8017.

Komhyr, W., Gammon, R., Harris, T., Waterman, L., Conway, T., Taylor, W., & Thoning, K. (1985). Global atmospheric co2 distribution and variations from 1968–1982 noaa/gmcc co2 flask sample data. *Journal of Geophysical Research: Atmospheres*, *90*(D3), 5567–5596.

Liu, J., Bowman, K. W., Schimel, D. S., Parazoo, N. C., Jiang, Z., Lee, M., … others (2017). Contrasting carbon cycle responses of the tropical continents to the 2015–2016 el niño. *Science*, *358*(6360), eaam5690.

Mao, J., Ramanathan, A., Abshire, J. B., Kawa, S. R., Riris, H., Allan, G. R., … others (2018). Measurement of atmospheric CO_2 column concentrations to cloud tops with a pulsed multi-wavelength airborne lidar. *Atmospheric Measurement Techniques*, *11*(1), 127–140.

Massie, S., Cronk, H., Merrelli, A., Schmidt, S., & Mauceri, S. (2022). Insights into 3d cloud radiative transfer for oco-2. *Atmospheric Measurement Techniques Discussions*, *2022*, 1–40.

Massie, S., Schmidt, S., Eldering, A., & Crisp, D. (2017). Observational evidence of 3-d cloud effects in oco-2 co2 retrievals. *Journal of Geophysical Research: Atmospheres*, *122*(13), 7064–7085.

Merrelli, A., Bennartz, R., O'Dell, C., & Taylor, T. (2015). Estimating bias in the oco-2 retrieval algorithm caused by 3-d radiation scattering from unresolved boundary layer clouds. *Atmospheric Measurement Techniques*, *8*(4), 1641–1656.

Miller, C., Crisp, D., DeCola, P., Olsen, S., Randerson, J., Michalak, A., … others (2007). Precision requirements for space-based data. *Journal of Geophysical*

Research: Atmospheres, *112*(D10).

OCO-2 Science Team, Gunson, M., & Eldering, A. (2022). *OCO-2 Level 1B calibrated, geolocated calibration spectra V11, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC).* https://disc.gsfc.nasa.gov/datacollection/OCO2_L1B_Calibration_11.html. NASA.

OCO-2/OCO-3 Science Team, Chatterjee, A., & Payne, V. (2022). *OCO-3 Level 2 bias-corrected XCO2 and other select fields from the full-physics retrieval aggregated as daily files, Retrospective processing v10.4r, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC).* https://doi.org/10.5067/970BCC4DHH24. NASA. doi: 10.5067/970BCC4DHH24

OCO-2/OCO-3 Science Team, Payne, V., & Chatterjee, A. (2022). *OCO-2 Level 2 bias-corrected XCO2 and other select fields from the full-physics retrieval aggregated as daily files.* https://doi.org/10.5067/8E4VLCK16O6Q. NASA. doi: 10.5067/8E4VLCK16O6Q

Parker, R. J., Webb, A., Boesch, H., Somkuti, P., Barrio Guillo, R., Di Noia, A., … others (2020). A decade of gosat proxy satellite ch¡ sub¿ 4¡/sub¿ observations. *Earth System Science Data*, *12*(4), 3383–3412.

Pasquarella, V. J., Brown, C. F., Czerwinski, W., & Rucklidge, W. J. (2023). Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 2124–2134).

Qu, Z., Jacob, D. J., Shen, L., Lu, X., Zhang, Y., Scarpelli, T. R., … others (2021). Global distribution of methane emissions: a comparative inverse analysis of observations from the tropomi and gosat satellite instruments. *Atmospheric Chemistry and Physics*, *21*(18), 14159–14175.

Ramanathan, A. K., Mao, J., Abshire, J. B., & Allan, G. R. (2015). Remote sensing measurements of the co2 mixing ratio in the planetary boundary layer using cloud slicing with airborne lidar. *Geophysical Research Letters*, *42*(6), 2055–2062.

Rayner, P., Law, R., O'Brien, D., Butler, T., & Dilley, A. (2002). Global observations of the carbon budget 3. initial assessment of the impact of satellite orbit, scan geometry, and cloud on measuring co2 from space. *Journal of Geophysical Research: Atmospheres*, *107*(D21), ACH–2.

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D., Avitabilie, V., Araza, A., ... others (2020). The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. *Earth System Science Data Discussions*, *2020*, 1–38.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., ... others (2020). The global methane budget 2000–2017. *Earth system science data*, *12*(3), 1561–1623.

Schepers, D., Guerlet, S., Butz, A., Landgraf, J., Frankenberg, C., Hasekamp, O., ... others (2012). Methane retrievals from greenhouse gases observing satellite (gosat) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms. *Journal of Geophysical Research: Atmospheres*, *117*(D10).

Stowe, L. L., Davis, P. A., & McClain, E. P. (1999). Scientific basis and initial evaluation of the clavr-1 global clear/cloud classification algorithm for the advanced very high resolution radiometer. *Journal of atmospheric and oceanic technology*, *16*(6), 656–681.

Tarrio, K., Tang, X., Masek, J. G., Claverie, M., Ju, J., Qiu, S., ... Woodcock, C. E. (2020). Comparison of cloud detection algorithms for sentinel-2 imagery. *Science of Remote Sensing*, *2*, 100010.

Taylor, T. E., Eldering, A., Merrelli, A., Kiel, M., Somkuti, P., Cheng, C., ... others (2020). Oco-3 early mission operations and initial (vearly) xco2 and sif retrievals. *Remote Sensing of Environment*, *251*, 112032.

Wilzewski, J. S., Roiger, A., Strandgren, J., Landgraf, J., Feist, D. G., Velazco, V. A., ... others (2020). Spectral sizing of a coarse-spectral-resolution satellite sensor for xco 2. *Atmospheric Measurement Techniques*, *13*(2), 731–745.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., ... others (2017). Comparisons of the orbiting carbon observatory-2 (oco-2) x co 2 measurements with tccon. *Atmospheric Measurement Techniques*, *10*(6), 2209–2238.

Yokota, T., Yoshida, Y., Eguchi, N., Ota, Y., Tanaka, T., Watanabe, H., & Maksyutov, S. (2009). Global concentrations of co2 and ch4 retrieved from gosat:

First preliminary results. *Sola*, *5*, 160–163.