Supporting Information for "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

Contents of this file

Introduction

Our overall analysis was performed in the major tropical areas but many of the figures in the main text only showed the Amazon. Here we provide a broader scope with a focus

---

on Africa and Southeast Asia as well. We also investigate potential optima in terms of footprint sizes. We will show a visual guide of our analysis, a brief investigation into the time-of-day dependency using OCO-3, as well as an evaluation of optimal footprint sizes.

Text S1: Visual representation.

Figure S1 provides a visual overview of our data analysis. Sentinel-2 images are acquired every 5 days, resulting in 6 snapshots per month. These are displayed as matrix in the left part of the figure, showing a 10x10km^2 area for clarity. For our analysis, we use the entire image size within 1x1 degrees. The left column shows Sentinel 2 reflectances as false-color image, with the reflectance at 835, 1610, and 2200 nm, respectively in red, green and blue. These bands are close to the OCO-2 bands, thus the red-ness of the scene indicated much higher O_2 A-band albedo compared to the CO_2 bands.

More than half of the scenes show typical shallow-cumulus cloud fields, which would greatly benefit from smaller ground-pixels for GHG measurements. Only five are mostly cloud-free, even for larger ground pixels than OCO-2 provides. A 2x2km area is indicated in blue in the first column. The visualization supports the skewed distribution observed in our statistical analysis. This results in long data-free periods during the wet season. The right column shows our corresponding binary cloud filter, i.e. black areas are cloud-free while white ones are either clouds or cloud shadows. Only one out of 72 images covering a 10x10km^2 area is entirely cloud free.

Text S2: Impact of time-of-day on clouds and data yields.

Figure S2 shows an analysis of the OCO-3 data yield in the Amazon, binning data yields within local overpass times. One can see that there are two maxima in the morning and afternoon when considering all OCO-3 provided in the pre-screened *lite* files. These peaks indicate an impact of the time of day on cloudiness and a possibility that the Sentinel-2 statistics is more benign (at 10:30 overpass time) than what can be expect for OCO-2 (13:30 overpass time). However, applying the stringent quality filter within the OCO-2 data eliminates these two peaks. It might have to do with the slant solar angles in these time-periods, which reduces overall observed radiances and increases the horizontal domain through which the direct solar beam propagates.

Regarding screening levels: OCO-2 (and OCO-3) retrievals often do not converge or have exceedingly high errors and these scenes are not considered further. The remaining fraction pass into what are called '*lite*' files, which are distributed to the scientific community. These nominally successful retrievals are further evaluated and classified as fully meeting the exacting requirements and therefore are assigned a quality flag = 0; lower quality retrievals are assigned quality flag = 1 or 2. Here, we define pre-filter yields based on the number of measurements that are provided in the *lite* files and the 'high quality' yields as those that also pass the quality flag = 0 within the *lite* files. In the main text, we only use high quality yields.

Text S3: Optimal footprint sizes.

In terms of effective revisit times, it is important to determine whether higher likelihood of obtaining valid data each time the satellite passes over the scene will outweigh the reduction in swath width using high-resolution measurements will. Provided we have a fixed across-track detector size and can only choose the footprint size x, the theoretical revisit time scales with $1/x$ and likelihoods to obtain measurements in a given scene scale with the median of our cloud-free statistics.

This allows us to derive the optimal revisit time per geographic location. Panel A in Figure S5 shows an example within the Amazon: while the cloud free fraction decreases consistently with footprint size, the effective revisit time exhibits an optimum footprint at which the revisit time is minimized. In this particular case, the 2 km footprint size would result in a 5 times longer effective revisit time than at 20-300m, while the theoretical revisit time at 2 km is almost 10 times shorter than at 200-300m. This underlines that both the data sparsity and revisit time problems in the tropics can be much better resolved with high spatial resolution rather than prioritizing as large a swath as possible. Panel B shows the map of the derived optimal footprint sizes across the humid tropics, for which the optimum is generally smaller than 500m. Surprisingly, the optimum outside the humid tropics quickly increase to >3km, indicating that the 1-2km resolution range might be too coarse for the humid tropics and too small for the extratropics. Given the importance of the tropics for GHG fluxes and the currently poor revisit times in these regions, prioritizing high resolution greenhouse gas measurements is needed.

References

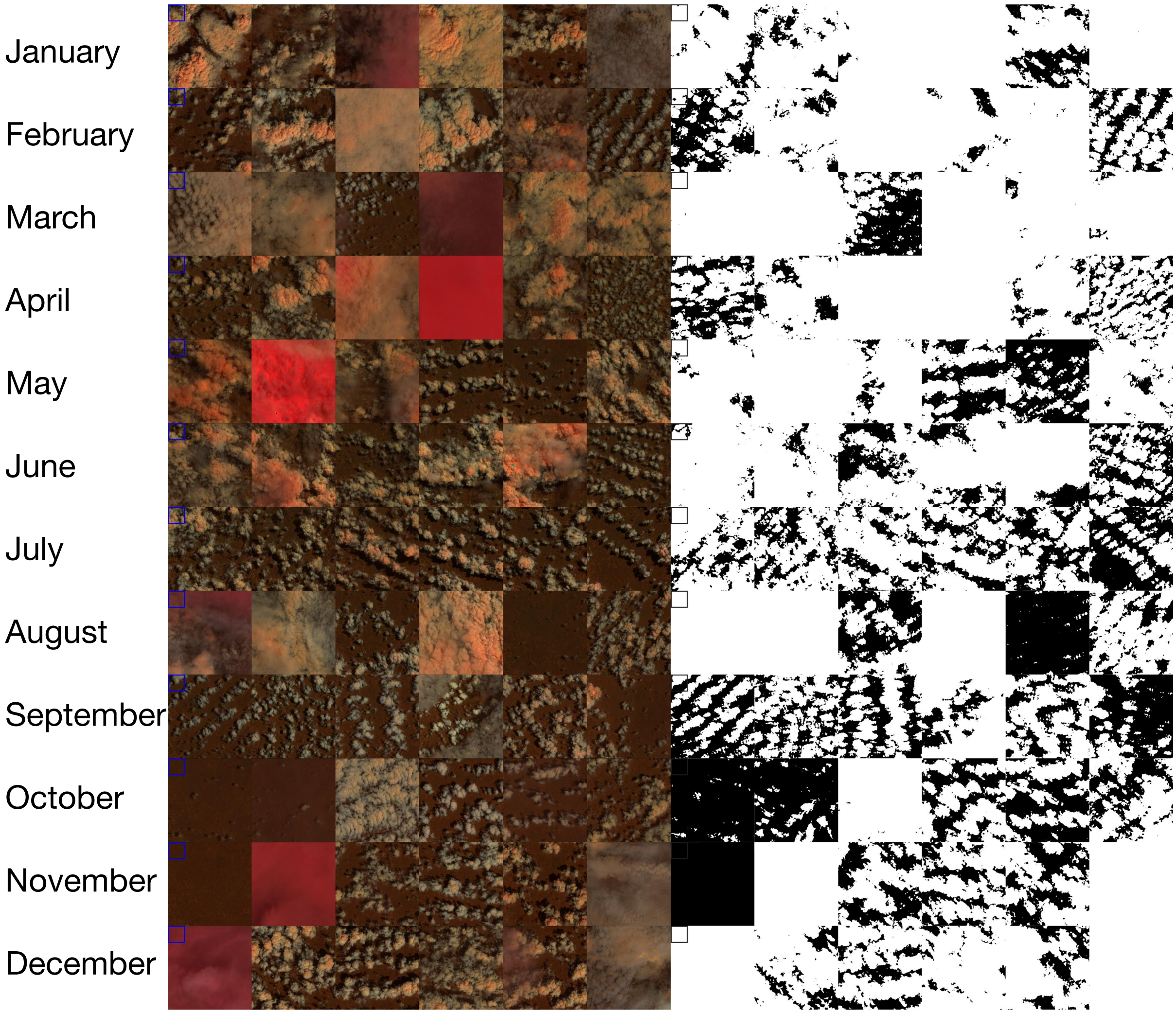

Figure S1. Visual representation of the Sentinel-2 cloud statistics using a full year top-of-atmosphere reflectance data in a 10km area (1.1-1.2N, 55-54.9W) in the Amazon, with 6 snapshots per months (5-day revisit cycle). The left columns shows a false-color image of reflectances, the right the binary cloud mask used in our analysis.

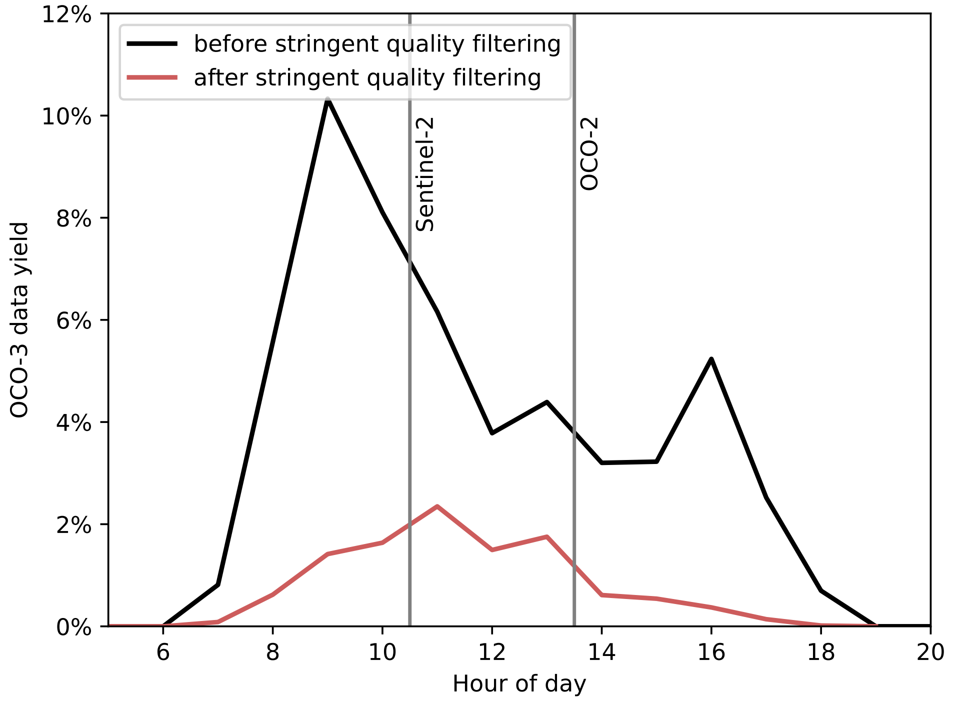

Figure S2. OCO-3 data yield in the Amazon as a function of time of day. The yields are compute with 1) all the data available in the XCO2 lite files (before quality filtering) and those within the lite file that pass the stringent quality filter (after quality filtering. While the loosely filtered data show some higher yields before noon and a secondary afternoon peak, the quality filtering largely suppresses these. Overall, there is no clear indication from OCO-3 data itself that shifting the local overpass time could greatly increase data yields by orders of magnitudes, which smaller ground pixels could achieve.

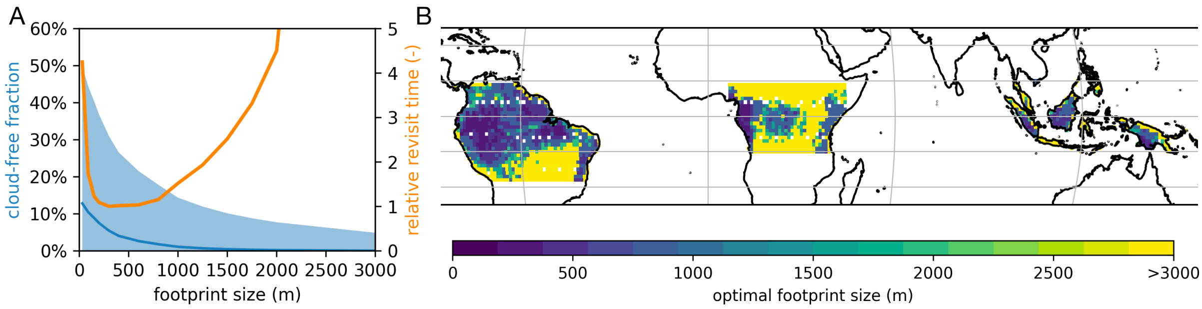

Figure S3. A) Median of the cloud free fraction (blue) and effective revisit time (orange, normalized by the minimum) as a function of footprint size for a larger domain in the Amazon (define). B) Map of the derived optimal footprint size that minimizes the effective revisit time. In most of the humid tropics, optimal footprint sizes are smaller than 500m, while they quickly grow to >3 km outside the areas that are most poorly sampled by current satellites.

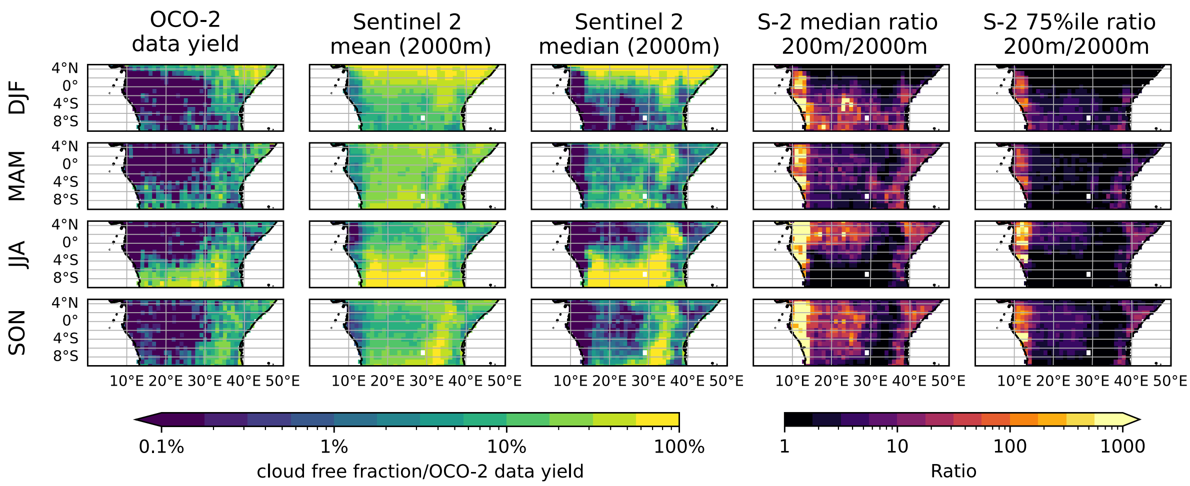

Figure S4. Same as Figure 2 but for tropical Africa.

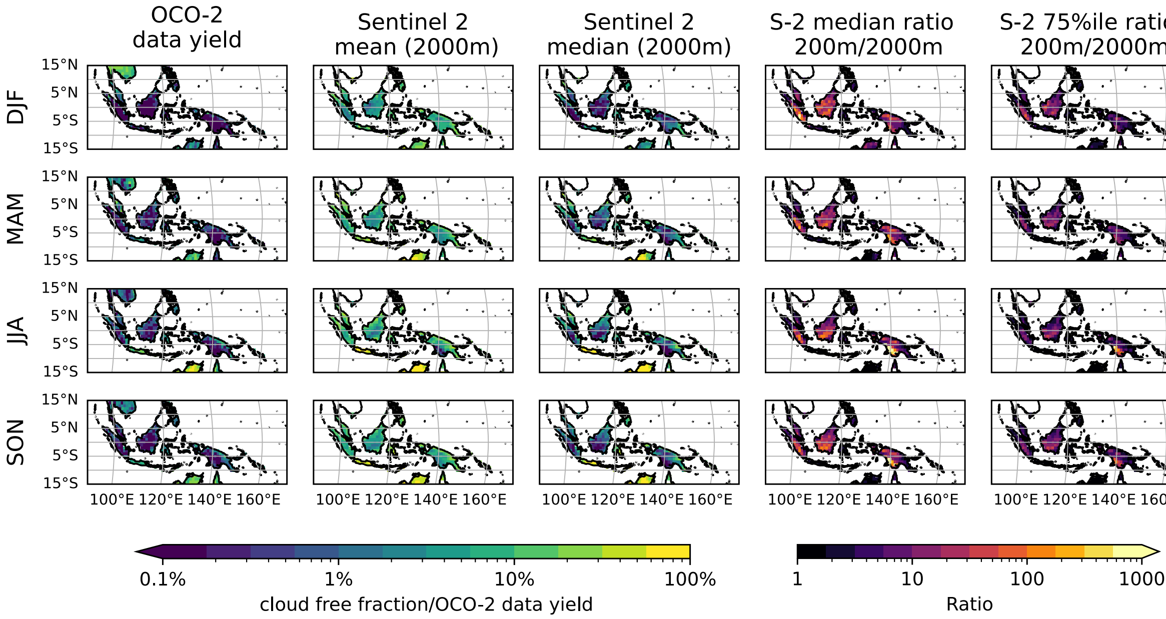

Figure S5. Same as Figure 2 but for tropical Southeast Asia.