General comments:

The manuscript entitled, “On the performance of satellite-based observations of CO2 in capturing the NOAA Carbon Tracker model and ground-based flask observations over Africa land mass” presents a scientifically interesting comparison of Carbon Tracker, GOSAT, OCO-2, and flask CO2 measurements. Despite Africa lacking ground-truth instruments such as TCCON, studies such as this one are useful for pointing out differences in the models and satellite observations. In general, there is one major methodological issue and many clarifications and technical fixes needed, but I recommend publication once they are resolved.

General comments:

- GOSAT and OCO-2’s primary product is the column-averaged dry-air mole fraction of CO2 (XCO2), not a vertical profile of CO2. There are typically less than 2 degrees of freedom for vertical CO2 for any given retrieval. Thus, the entire comparison to flasks should come with a disclaimer that the NASA L2 retrievals for GOSAT and OCO-2 are not designed to be used in this way. The comparison is still interesting, but I am unsure about the scientific value.

- The authors often list characteristics of a certain region (e.g. high anthropogenic emissions, low vegetation levels) and then attribute the difference between CT and GOSAT/OCO-2 to these characteristics. The data is indicating correlation, not causation. Additional research (e.g. a detailed modelling study) would need to be done to provide evidence that the XCO2 difference is *caused* by such characteristics. I note several instances of this below where it would be wise to soften the language.

- For all the maps, I would strongly suggest not to use the default rainbow colormap for XCO2. Depending on the coding language you use, there are a number of much better colormaps available. For ordered information, such as XCO2, you should use a perceptually uniform colormap (such as viridis in Python). For diverging data, such as CT2016 – GOSAT, you should use a diverging colormap (such as RdBu in Python) and center the colorbar at 0. In many of your figures, you use a rainbow colormap with unequal positive and negative limits, which makes it incredibly difficult to determine where on the map the bias is above or below zero.

https://matplotlib.org/tutorials/colors/colormaps.html

- When discussing the distance between a given GOSAT/OCO-2 measurement and CT, could you please elaborate on what exactly this means? Each GOSAT/OCO-2 measurement should fall within a CT grid cell, so dx seems meaningless to me.

- The mean bias for the entirety of Africa is mentioned numerous times, including in the abstract. However, your analysis shows that there are large regional patterns. Thus, there is little scientific value in, for example, stating that GOSAT XCO2 is 0.28 ppm higher than CT. Additionally, no uncertainties are given for any statistics in this paper. This should be resolved before publication. For example, 0.28 +/- 1.5 ppm is much less meaningful than 0.28 +/- 0.2 ppm.

- For OCO-2, are you using land nadir data, land glint data, or both? For GOSAT, you are presumably including the medium gain data, but please state so.

Specific comments:

- P2 L30: Citation for this? The land surface characteristics could affect retrievals, but I'm unaware of the impact of anthropogenic sources on satellite XCO2 biases.

- P3 L9: This makes it sound as if models are intrinsically more accurate than the satellite measurements. If this were true, why would we even need satellite measurements? In general, however, the paper does a good job at saying the models and obs. "agree" or "disagree" rather than one is "wrong" or "right."

- P4 L10: SCIAMACY measured CO2 and CH4 before GOSAT.

- P4 L19: GOSAT ACOS B3.5 is now ~5.5 years out of date. B7.3, which represents a significant update to the retrieval, has been available for over 3 years now. It is too much to ask of the authors to repeat their analysis with the newer version, but it must be noted that the version used is very outdated. See the official Data Users Guide for details on the latest product:

https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/ACOS_v7.3_DataUsersGuide-RevF.pdf

- P4 L26: Please cite some OCO-2 papers in this section (e.g. Crisp et al., 2008,

- P5 L16: If CT is a 3-hourly product, the maximum d(time) would be 1.5 hours.

- P7 L10: Citation needed regarding Southern Africa's characterization.

- P7 L11: How do you know that this is the reason for the bias dipole?

- P7 L19: How would low number statistics result in a high bias? It's certainly possible, but no explanation or mechanism is provided.

- P7 L19: Citation needed regarding rainfall.

- P8 L1: These plots are very difficult to interpret because of the large number of data points. I would strongly suggest to instead plot heatmaps of the XCO2 difference vs. the spatial difference. And, as noted above, it is not clear what the distance metric actually represents.

- P9 L5: The higher GOSAT/OCO-2 uncertainty in these regions is likely driven by low signal to noise in the strong CO2 band over dark forests.

- P10 L6: Could use a general citation here.

- P12 L15: If the CO2 sink is growing after the rainy season, why would GOSAT not see it?

- P14 L1: Same as above: why would there be a difference? You seem to imply that the difference must be because of local sources and transport, yet this is speculation. I would simply soften the language from “likely” to “possibly.”

- P17 L4: The cirrus cloud hypothesis should be removed unless you can show that there are more cirrus clouds over that specific region which could potentially be biasing the satellite results.

- P17 L11: By what mechanism would a cold bias impact the CT XCO2? Would suggest removing unless you can provide a reasonable hypothesis.

- P17 L18: How would low vegetation levels and local sources result in a low correlation between the two products? Would suggest removing unless you can provide a reasonable hypothesis.

- P19 L17: Good. Here, a correlation is discussed (higher OCO-2 where there’s more vegetation) without asserting causation. Another hypothesis could be cloud contamination in the satellite retrievals.

- P23 L9: What plantation is this referring to? Please elaborate or remove this statement.

- P25 L11: What intensive fire is this referring to? Please elaborate or remove this statement.

- P29 L2: This is a disappointingly brief discussion on reasons why the model could have issues. This paper should emphasize that neither models nor satellites are perfect, and that all that can be done in a poorly constrained place such as Africa is a comparison and discussion of potential reasons for the differences. For example, clouds, aerosols, and dark surfaces can result in biased XCO2 from satellites, while poor parameterizations and insufficient input data can hinder models.

- P29 L4: Should thank both the appropriate Japanese agencies for GOSAT and NASA JPL for the GOSAT ACOS and OCO-2 retrievals.

Technical comments:

There are numerous spelling and grammar issues that should not be the responsibility of a reviewer to fix. I would suggest that the authors spend some time resolving these issues.

Overall: XCO_2 is never defined.

- P3 L25: “combines observed in situ carbon dioxide”
- P7 L15: Likely a typo. GOSAT in comparison to GOSAT.
- P10 L2: Oddly worded. Just say Africa has significant land mass in both hemispheres.
- P27 L17: Oddly worded. Perhaps, “is important to identify differences between GOSAT and CT.”

Figure comments:

- As stated above, please use appropriate colormaps and colorbar ranges for diverging data.
- For time series, please use years and months instead of “months since.”