The authors provide measurements of methane emissions from 21 offshore facilities using two different aircraft-based platforms. The findings basically show the utility of such measurements to validate national emission inventories and improve emission estimation practices, including national reporting guidelines. The main implication that the authors communicate is the need for more measurements; however, the results can also be used to help understand the root cause of the emissions and to help improve ways in which the industry reports their emissions and operations. Overall, the paper is well-written and only needs minor modifications.

Although the paper does a very nice job of providing information that can validate inventories, I think the authors can provide more information on characteristics of the offshore facilities that can be used to identify the root cause of the emissions. For example, even a simple differentiation of facilities by the type of hydrocarbon being produced could be a helpful start. Moreover, I think the paper points to some opportunities in which emissions can be reported better by industry and if so, these points should be made more explicit. Below are some detailed suggestions.

Line 30: It would be good to add the word “only” in front of “16%” to emphasize that 16% is small.

Line 38: What is meant by “this”? Is “this” “measurements of temporal variations” or “knowledge of facility operational status over time”?

Line 39: How big of a sample is sufficiently large? Knowing how many times a facility should be measured would help regulators. How do we know we have a representative sample?

Line 40: Specify “aircraft” in front of “measurement approaches”, unless the point here is that operators use any measurement approach.
Lines 56-57: How much of the methane emissions from the O&G sector does offshore O&G production represent? This type of information would be valuable somewhere in this paragraph.

Lines 63-65: Are these studies offshore or onshore studies? How do the authors expect the offshore emissions to be different from onshore emissions?

Line 73: What is the ACCESS campaign? Spell out ACCESS.

Line 103-104: A short sentence here on how the FAAM platform is much larger than the Scientific Aviation platform would be helpful here. Also, add “two” in front of “aircraft platforms”.

Figure 1: It would be helpful if the oil and gas production was presented in such a way that we can distinguish between oil and gas (and even condensate or mixed), in addition to production amounts.

Line 188: The production numbers in Figure 1 are 3-4 orders of magnitude lower compared to what’s in Figure 2. Is this a typo? And importantly, are these production amounts really large? What are the amounts of oil and gas produced in other regions (e.g., Gulf of Mexico)?

Line 198-199: Were the operators broadly aware of the measurement study, even though they did not know when the measurements were happening?

Figure 2: These oil & gas production numbers are 3-4 orders of magnitude larger than those in Figure 1.

Figure 2: Could the facility numbers be shown here as well? Also, what is the difference between the cluster of red facilities and the rest, which are mainly blue? It seems that there is clearly a spatial pattern, possibly governed by geology and the produced hydrocarbon (i.e., gas vs. oil).

Line 261-262: How do the Norwegian guidelines compare with the IPCC guidelines and guidelines from other countries (e.g., U.S.)?

Line 268-269: Sentence is not complete.

Line 272: A list of all potential sources would be helpful here, especially for mitigation.

Equation 1: Define A and B and zmax.

Line 327-328: How many vertical layers are considered at the sites? I recognize that this will be a range.

Line 333-335: How different are the mixing latitudes obtained using these different approaches?

Line 397: Do the Norwegian guidelines require reporting of emissions at the annual level? In such annual estimates, there would have been estimates of the number of emission events per year.

Line 438: I'm not sure if it makes sense to say these measurements are "outliers". They are still a part of operations (i.e., turbine maintenance). Instead, I think it points to how the inventory guidelines could be improved to be more consistent with operations.
Figure 4: It would be good if these plots could indicate the operation or activity that the emissions are coming from. Above, for the outliers, cold-venting and turbine maintenance are mentioned. But what is happening at the other measured facilities?

Figure 4: Could we get more information on the facilities? For example, are they mainly producing oil or gas?

Line 470: How do you know the data points in the blue boxes are outliers? How are "outliers" defined? Visually, the left data point (around 300 t/year) doesn't really look an outlier to me.

Line 483: What is the difference between Facilities 6 and 7? Could this large difference have been predicted?

Line 494: Would the facilities measured here be considered deep or shallow water facilities?

Line 504: Which inventory for 2019? The scaled Scarpelli et al inventory? If so, please specify.

Line 557: Is it reasonable to assume that the errors are normally distributed? Also, are the sample really independent? There are likely to some similarities among facilities with the same operator and in the same basin. And there are likely other relevant factors such as geology, production, and technology.

Line 564: a missing "to" in front of "lead"

Line 564: Can we say something about how many times per year a facility should be sampled?

Line 581-582: There may be a need to reconsider the use of the term "normal operations". Many of the "other" operations such as maintenance are a part of the normal activities on a platform and may be better viewed as routine.