**Reviewer 1**

**Major Comment:**

I appreciate the efforts from the authors. The revised manuscript has addressed some of my previous concerns. However, I am now more concerned with the inconsistency between the case study and the statistics in this revised manuscript. According to the overall statistics in Fig. 6, Vmax in CV31 is degraded in the first 12 hours. Even for the single Ian case in Fig. 7, CV31’s Vmax is not apparently improved. Despite this, in the two case studies, the authors attribute the overall improvements in CV31 to the improved analysis, particularly from the surface wind field and the derived fields at the air-sea interface. The two cases showed improved Vmax predictions for at least the first 24 hours, which does not align well with the statistics. It seems these are cherry-picked cases where CV31 performs better, while there may be other cases where CV31 performs worse initially; otherwise, the significant improvements in Fig. 12b should be reflected in Fig. 6b/7.

In summary, a case study should ideally illustrate the features from the statistical results and provide a reasonable explanation for the improvements or degradations observed in the statistics. However, in this study, it appears the statistical results and the case studies are presenting two different narratives. I do not doubt that the hypotheses from the case studies suggest potential benefits of assimilating the CyGNSS L2 wind data, but I question whether they are the sole reason why CV31 statistically performs better overall. I recommend that the authors link the case study results more closely to the overall statistics or at least provide an explanation for why other cases showed worse initial Vmax in CV31, yet it catches up in the long-term forecast. And since Pmin is improved more at the analysis time, perhaps digging deeper into the pressure side (e.g., wind-pressure relation) could help.

[Lew: Added bad case figure for Ian in SI; add text to Results, Discussion, Abstract to clarify that results are mixed.]

[Bachir TBD: Add bad case figure(s) for Earl[? Or Fiona?] in SI, add text to Results, Discussion, Abstract to clarify that results are mixed.]

Minor comments:

1. L340, 351: Please note they are “the overall track results”, and “the overall Pmin and Vmax results”

2. L342: 4.7 %

3. L353: only “eight of 22”?

4. Fig. 8b: related to my major comment, this figure indicates comparable initial Vmax between CV31 and CNTL. Yet in your response figure R1-3, the CNTL appears to have a much stronger wind fields than CV31. Do you have any comments on this inconsistency?

Fig 8b shows the Vmax analysis at the surface 10 m. Fig

5. L594: I think the RMW in Fig. 6 has been removed.

**Reviewer 2**

The authors have come some way in my direction, but I feel the effort was insufficient. The review process – and in particular the responses from the authors – has allowed me to get to the bottom of some of the issues that bothered me in the first place. I think this can be published, but it’s going to require some additional major changes and also more serious consideration of points that I previously raised. Please carefully consider my points below and make changes of substance.

**Major comments:**

1. This paper would be much better if it focused only on Hurricane Ian, for which you have analyzed the entire storm. As it currently stands, it feels arbitrary that you included a handful of cycles from a number of TCs and then also the entirety of Ian. Adding to the problem, the entire analysis of Larry misses the point of my prior round of review. The cycle analyzed has a major spindown issue in CTRL that has nothing to do with the ocean – the ocean simply responds to the atmosphere. That entire section needs to be removed. On the other hand, there are some potentially compelling arguments to be made about the ocean in some forecasts of Ian.

2. You haven’t included any figures that show when/where the additional data is assimilated – that information is available in the GSI diagnostic files. In a paper that focuses only on Ian, you could actually include a time series by cycle of the amount of additional data assimilated as well as a figure that shows a few examples of where in the storm the data gets in.

[Bachir: Add to Supplemental Info this figure.]

3. Figures 4, 8, 9, 11, 12 – much of the text in these figures is still too small to read. Please do not include text in a figure if it cannot be read without substantially zooming in. I raised this issue in the previous round of review.

[Bachir: redo 4 as a single panel? Lew: crop titles from Fig 8c/d, 9, 11b/d, 12c/d]

4. The discussion on lines 382-390 is really not formulated well. What I see in Fig. 7 is general improvements in many aspects of the forecast in the first 24 h, but after 24 h I see only slight improvements in the track and R34 forecasts. Everything else is a mixed bag, and highlighting where only a few isolated lead times have improvement is distracting and within the noise. Please consolidate and focus this discussion on the points that most matter. One other thing – on line 386 the improvement is described as “moderate”. It’s not clear what “moderate” means here. In the consistency metric, MCI means “marginally consistent improvement”.

[Bachir: Add storm name and sample size to Fig. 7. Change “moderate” to “marginally consistent improvement”.]

5. Lines 466-478. A few things here. I think what you’re implicitly saying is that the positive change in height implies convergence, but that is never stated. Why not just show convergence/downwelling instead of going the roundabout way? Do you not have the ability to perform those diagnostics in the ocean model you use?

[Lew: Add to SI: Ocean surface convergence figures.]

6. There are serious problems with Fig. 11 and commensurate arguments in the text. First and foremost, it’s from an entirely different forecast from the rest of the preceding analysis! That’s extremely misleading to the reader. Second, the panels are not labeled consistently with the text (panels b and c in particular. Third, I can’t make sense of the SSH plots in comparison with the text. In particular, lines 483-485 claim a greater extent of SSH > 1 m in CTRL, but I sure can’t tell that from the figure. In fact, the CV31 SSH looks higher AND CV31 has a stronger TC. This is exactly opposite of what Fig. 8 shows. This begs the question – are the tendencies outlined for the 27/18 cycle consistent with other cycles?

[Lew: Fix Fig. 11 and text to show 0927/18.]

7. Lines 491-505. It’s hard to tell what the point of this paragraph is, and there are conflicting statements. For example, the authors dropped this statement: “The authors clearly acknowledge however, that the anomalous intensification of the CTRL in the very early forecast (hours 3-12) would have been driven largely by other differences in the near-storm environment between the experiments, beyond the scope of the present analysis” in response to my previous comments. Well fine, but that contradicts the rest of the paragraph. So what’s the point? Is it that both the initial atmospheric conditions and ocean response contributed to the stronger TC in CTRL? If so then that needs to be stated clearly and up front. But you also need to prove the part about the ocean with a cleaner assessment from the appropriate forecast cycle.

[Lew: Add to Fig. 10 and 13 the CYCLE date/hour. Add text to both paragraphs as suggested above.]

Minor comments:

1. Lines 138-140: Not clear what this means

[Bachir: rewrite those sentences]

2. Figure 2 – there’s no scale

[Bachir: add legend]

3. Lines 253-254: The observations are listed here in a way that doesn’t entirely make sense. Please list as: “conventional (including terrestrial, aircraft, and satellite datasets), clear-sky satellite radiances, ground-based radar, METAR, and reconnaissance data.”

[Bachir: good idea above.]

4. Line 254 – Zhan et al should be Zhang et al

[Bachir]

5. Lines 327-329 – unclear what the point is

[Bachir]

6. Figure 5 – there’s a stray “ab” on the figure

[Bachir]

7. Figure 6 – there’s a stray “a” on the figure

[Bachir]

8. Lines 351-357: The improvement in PMIN is not particularly consistent as judged by the consistency metric even though the MAE improves. This suggests that assimilating the additional data reduces the number of PMIN outlier errors but does not improve results consistently across the sample.

[Bachir: rewrite to be consistent]