We express our sincere appreciation for the insightful and constructive review of our manuscript. Our detailed, point-by-point responses to the specific comments provided are as follows:

1. It would be worth mentioning some aspects of the WRF/WRFDA model configuration that differ from operational global and regional TC forecasting models. For example, the 3DVar algorithm uses just a static background error covariance matrix without ensemble-based "errors of the day." Also, impacts of the relatively coarse 15-km horizontal resolution: Might this explain why some cases experienced TC genesis but quickly dissipated, such as Gaston in the 14kBestRO experiment? I realize that this resolution is not much coarser than current operational global models that the forecasters use for TC cyclogenesis prediction. But if global forecast model horizontal resolution continues to increase in the future, might we expect to see greater value in utilizing RO data for cyclogenesis prediction, and how would this study's results apply to those future models?

We agree that our current WRF/WRFDA configuration, which utilizes parameterized convection and a static 3DVar background error covariance, is not optimal for TC genesis prediction. We acknowledge that increasing horizontal resolution to allow explicit prediction of convection will reduce uncertainties associated with convective parameterization. Such improvement could potentially prevent premature dissipation events such as that observed in the case of Gaston. Furthermore, the hybrid 3D-EnVar (which includes both static and ensemble-based error covariances) would assimilate the RO data more effectively.

Regarding the future value of RO data in high-resolution models, we believe the impact will depend heavily on its spatiotemporal distribution. While increased model resolution is essential, the information content is constrained by the spatial density of the observations.

Ultimately, the core finding of our study that the spatiotemporal distribution of RO data influences TC genesis prediction is clear. We expect this conclusion to remain valid even in more advanced, high-resolution, ensemble-based modeling frameworks, though we agree that additional experiments using advanced data assimilation methods with a higher resolution would be necessary to quantify these benefits.

We will incorporate this discussion into the Summary section of the revised manuscript.

2. Could you please provide some information regarding the prescribed RO refractivity observation errors? Do they vary according to satellite mission, latitude, altitude, or atmospheric conditions?

The vertical profiles for prescribed RO refractivity observation errors, adapted from the nonlocal operator framework of Chen et al. (2020), are presented in Figure R1. Currently, these error specifications are purely altitude-dependent and do not account for variations across different satellite missions, latitude, or specific atmospheric conditions. It would be interesting to explore these variations in future studies.

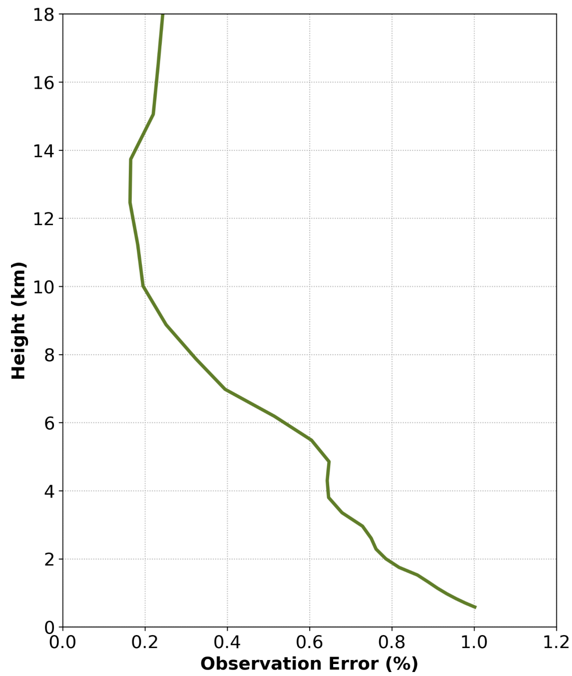

Figure R1: Vertical profile of the prescribed refractivity observation errors used in this study. Adapted from Chen et al. (2020).

3. Have you explored sensitivity of your results to variations in the vortex tracker settings, such as the 50-km radius used for defining the circulation?

We appreciate this suggestion. While we had not previously explored the sensitivity of our results to the vortex tracker settings, we have since tested the impact of varying the circulation radius by adjusting it to 30 km and 100 km (Figure R2). While a 30-km radius increases the number of vortex candidates identified within the model domain, the center location of the developing TC and the simulated genesis time remain consistent across these settings, indicating that our results are robust.

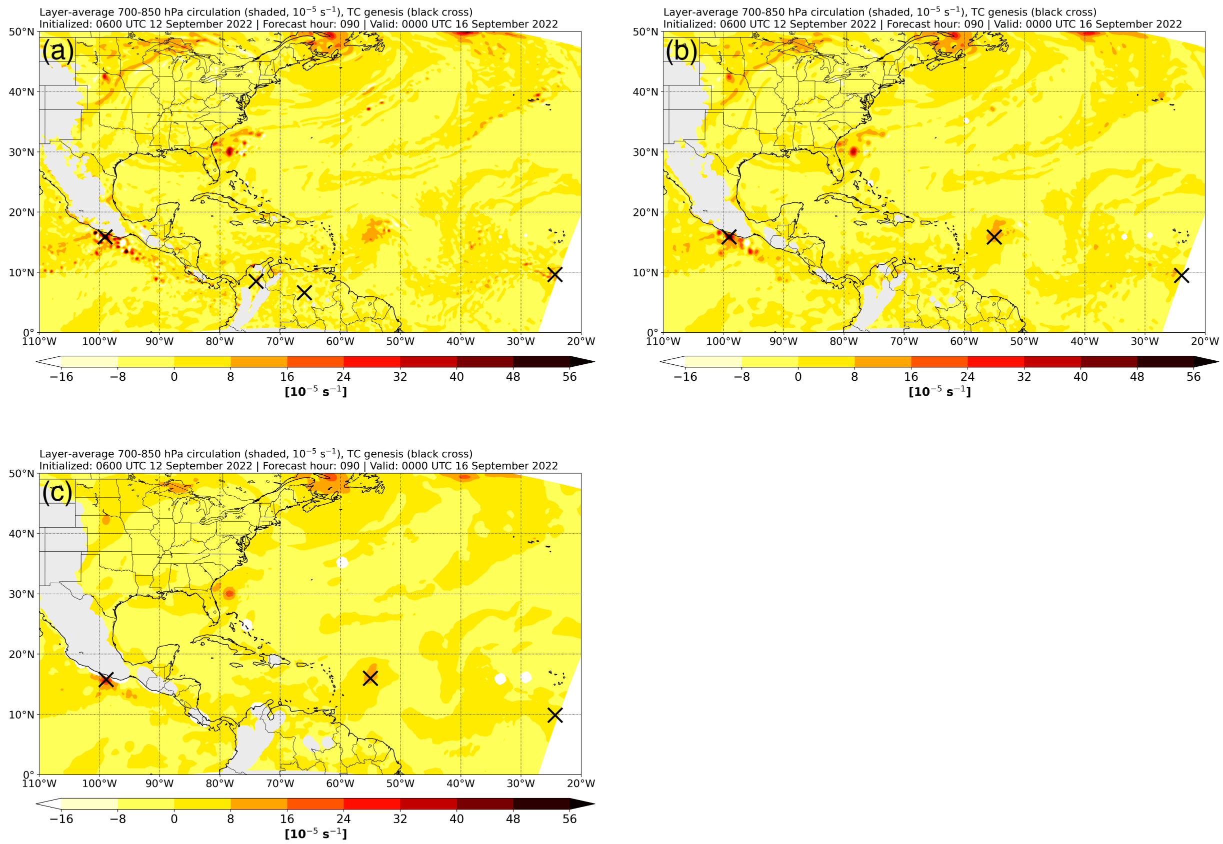

Figure R2: 700-850 hPa layer-averaged circulation (shaded; 10^{-5} s^{-1}) averaged within (a) a 30 km radius, (b) a 50 km radius, and (c) a 100 km radius of a grid point from the 14kBestRO experiment at 0000 UTC 16 September 2022. The black cross marks the disturbance that meets all criteria of the TC tracking algorithm.

4. Probability of Detection statistics shown in Table 2. It appears that you are not counting the events where genesis briefly occurred, but the TC quickly dissipated, shown as triangles in Table 2, in the POD values listed for each experiment. For example, the POD for 8kRO is listed as 1/6, which would, among all cases, just be counting Julia as a hit within the +/- 24-h time tolerance criterion. Therefore, the triangle for Fiona must not be counted, even though in lines 180-181 you say that such cases are counted as a "0.5 hit". Similarly, the 14kBestRO experiment POD is listed as 2/6, which presumably comes from Julia and Nicole being successful hits, but the 0.5 hit from Gaston is not being counted here. Maybe I'm missing something?

Thank you for your careful review and for pointing out the inconsistency regarding the probability of detection (POD) calculations in Table 2. We agree that our previous description was confusing; the term '0.5 hit' was misleading and did not align with our quantitative results.

To clarify, we have reviewed our methodology to strictly define a ‘hit’ as a case where the model-predicted genesis time falls within the ± 24-h window of the observed time (following that of Halperin et al 2013). The ‘triangle’ cases (e.g., Gaston in the 14kBestRO experiment) are identified as non-intensifying disturbances; consequently, they are explicitly excluded from the POD calculation. We will remove the ‘0.5 hit’ terminology from the manuscript and update the text to ensure complete consistency between our definition of a ‘hit’ and the values presented in Table 2.

5. Figure 9 and related discussion of the Hurricane Fiona RH and vorticity fields at t=48 h (lines 284-294). Even though 14kBestRO shows the “strongest and broadest moistening” among the three experiments, as you write in line 286, there are some notable differences between this experiment’s 48-h forecast RH/vorticity fields and the GFS analysis shown in Figs. 9a, 9b. For example, whereas the GFS analysis shows 75% RH extending above 350 hPa (Fig. 9a), the 14kBestRO experiment’s 75% RH contour extends only up to around 500 hPa, with very dry air in the upper troposphere all the way to the TC center - more so than in the other two experiments. Additionally, even though you point out the difference between the “very narrow vorticity structure of 14kPoorRO” compared to the GFS (line 291), it’s worth noting that despite its being confined within the 0.5-degree radius, 14kPoorRO’s azimuthally-averaged vorticity profile (Fig. 9h) more closely resembles that of the GFS (Fig. 9b) in terms of its greater depth, compared to the other two experiments, which forecast a shallow vortex that extends up to around 500 hPa (Figs. 9f,9g). Could it be possible that the GFS and 14kPoorRO are correctly capturing a tighter and deeper circulation for Fiona at this time and that 8kRO and 14kBestRO are failing to capture the deeper vortex, despite their doing a better job than 14kPoorRO in capturing the wider vortex closer to that of the GFS in the lower troposphere? Could you provide some additional comment on these structural differences shown in Figure 9 that seemingly contradict your main point about 14kBestRO forecasting inner-core RH fields and vortex structure for Fiona that are closest to the GFS analysis?

We thank the reviewer for the careful examination of Fig. 9 and for pointing out apparent inconsistency between the forecast vortex structures and the GFS analysis. Upon re-examining the diagnostics, we identified an issue in the computation of the azimuthal averages used in the original figure. The original azimuthal averages were computed over the area enclosed within each radius, which artificially influenced the apparent vertical structure of the vortex. We have corrected the diagnostic by computing azimuthal averages along each concentric radius ring and have changed the horizontal axis from radius (degrees) to distance from the vortex center. Figure 9, as well as Figure 8, have been updated accordingly.

With the revised diagnostics, the 14kPoorRO experiment no longer exhibits the deeper upper-level vortex suggested by the original figure. Instead, it shows substantially weaker upper-level vorticity (approximately 200-400 hPa) and a drier inner-core environment than both the GFS analysis and the 14kBestRO experiment. We will revise the corresponding discussion to reflect these updated results.

We also note that the revised figure still indicates that none of the experiments perfectly reproduces every aspect of the analyzed vortex structure, particularly the upper-tropospheric moisture distribution.

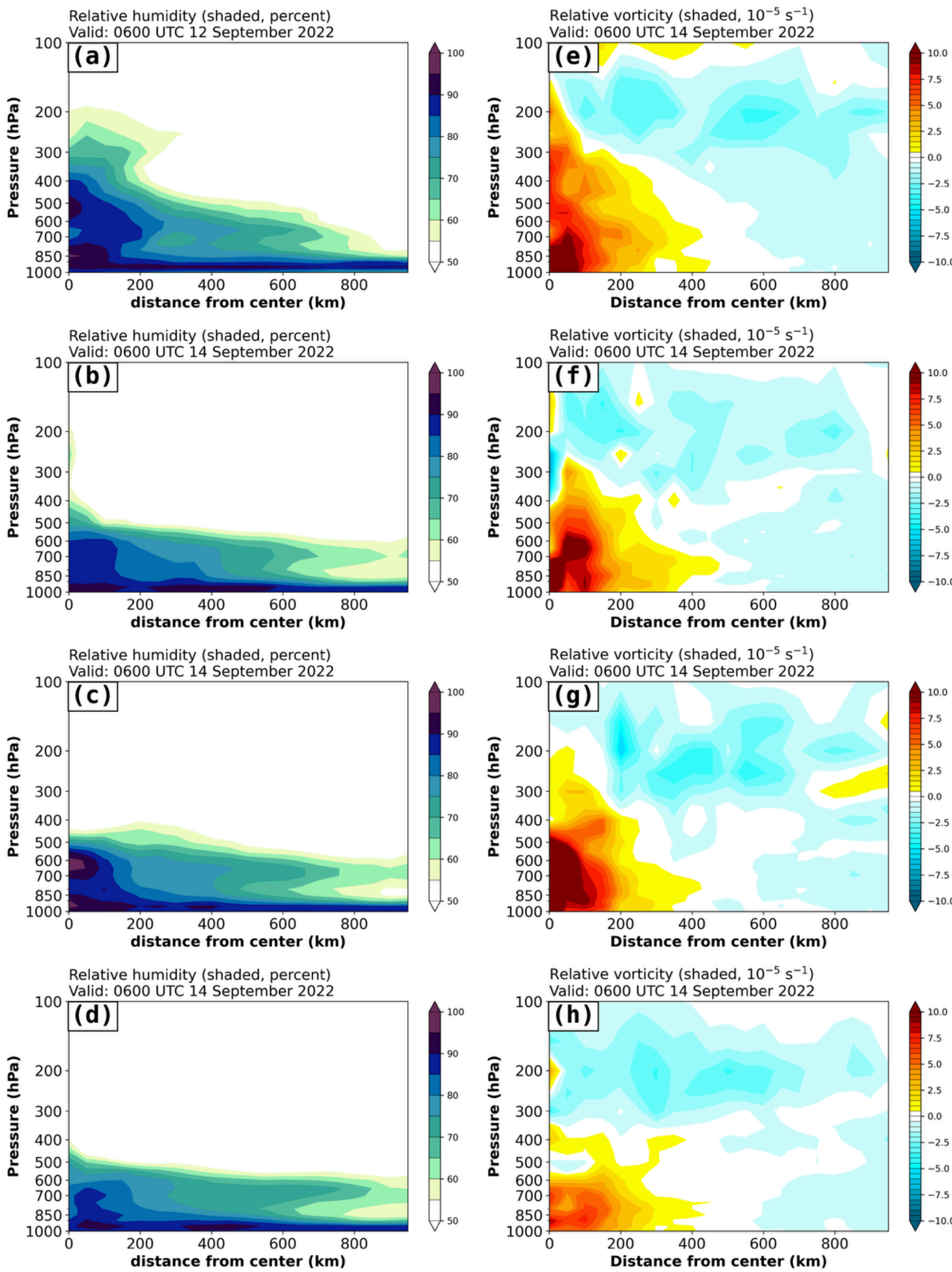

Figure 9. Radius-pressure cross-sections of (left column) relative humidity (%) and (right column) relative vorticity (10-5 s-1) for (a, e) the GFS analysis, (b, f) 8kRO, (c, g) 14kBestRO, and (d, h) 14kPoorRO experiments at 0600 UTC 14 September 2022.

6. When describing the failed Ian genesis case forecast by 14BestRO, you write in lines 340-341: "However, the modest moisture increase was insufficient to counteract the

ventilation effects of vertical wind shear." Do you have evidence showing that the drier vortex shown in the three simulations, compared to the GFS analysis, was due to vertical wind shear-induced ventilation effects, as opposed to a dry bias in the WRF analysis for the eastern Caribbean region (Fig. 13; perhaps due to model physics or the DA algorithm) that could not be corrected enough in any of the simulations to enable cyclogenesis to happen? One way to demonstrate the vertical wind shear-induced ventilation effects could be to compare the radial wind and moisture anomaly covariances for Ian's vortex between the GFS analysis and the three experiments at different levels and compute their environmental vertical wind shear vectors. Otherwise, it might be best to rephrase this sentence to suggest that ventilation effects are just one possible explanation for the drier WRF forecast Ian vortices, compared to the GFS.

Thank you for this constructive suggestion. To address your concern regarding whether the drier vortex was driven by shear-induced ventilation or a dry model bias, we followed Tang et al. (2024) to compute the covariance between radial wind anomaly and equivalent temperature anomaly within a 500-km radius of the storm center at Hurricane Ian's genesis time (Figure R4). Our analysis reveals a strong negative covariance between 850 hPa and 600 hPa in the 14kPoorRO and 8kRO experiments, which is indicative of ventilation effects. In contrast, the 14kBestRO experiment and the GFS analysis show positive covariance in these layers. These results support our hypothesis that ventilation, rather than a broad dry bias in the WRF analysis alone, was a primary factor limiting the vortex intensification in the DA experiments. We will revise the manuscript to acknowledge that while our diagnostic analysis points to shear-induced ventilation, the dry bias of the WRF analysis remains a potential contributing factor in the development of the vortex structure.

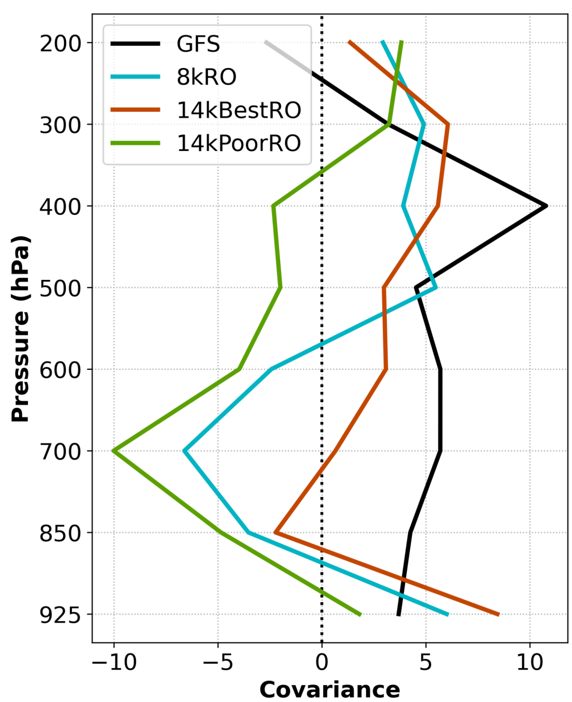

Figure R4: Vertical profile of radial wind and equivalent potential temperature anomalies within a 500-km radius from the storm center at 0600 UTC 23 September 2022 for the GFS analysis (black), 8kRO (blue), 14kBestRO (red), and 14kPoorRO (green) experiments.

Reference:

Tang, B. H., R. Rios-Berrios, and J. A. Zhang, 2024: Diagnosing radial ventilation in dropsonde observations of Hurricane Sam (2021). Mon. Wea. Rev., 152, 1725-1739, https://doi.org/10.1175/MWR-D-23-0224.1.

Halperin, D. J., H. E. Fuelberg, R. E. Hart, J. H. Cossuth, P. Sura, and R. J. Pasch, 2013: An Evaluation of Tropical Cyclone Genesis Forecasts from Global Numerical Models. Wea. Forecasting, 28, 1423 - 1445, https://doi.org/10.1175/WAF-D-13-00008.1.

Minor/editorial comments:

Line 50: Recommend rephrasing “GNSS RO observations can more accurately capture the thermodynamic environment which are favorable for the development” to “GNSS RO observations can more accurately capture thermodynamic environments favorable for the development” or something similar

Thank you for the suggestion. We will rephrase it in the revised manuscript.

Line 95: “The GFS analyses is” —> “The GFS analyses are”

Thank you for the suggestion. We will revise the sentence in the revised manuscript.

Lines 96-97: Probably better to rephrase “the best track data from the National Hurricane Center (NHC), National Oceanic and Atmospheric Administration (NOAA)” to “the best track data from the National Oceanic and Atmospheric Administration (NOAA)’s National Hurricane Center (NHC)"

Thank you for the suggestion. We will rephrase it in the revised manuscript.

Line 103: “Individual DA experiment is” --> “Individual DA experiments are”

Thank you for the suggestion. We will update it in the revised manuscript.

Lines 204-205: In the Figure 5 caption, the first sentence is long and cumbersome. Recommend placing a period after “14kPoorRO (orange) experiments”. Then for the next sentence, use something like “The left panels (a, d, g) show the average among all cases, middle panels (b, e, h) show results for Hurricane Fiona, and right panels (c, f, i) show results for Hurricane Ian.

Thank you for the suggestion. We will update the caption in the revised manuscript.

Lines 211-217: For the RMSE analysis shown in Figure 7, are the data points collected for generating RMSE statistics collected from the full 120-h free forecast period? Are data also collected from the DA cycling period? It would be worth clarifying this in the text.

The data points collected for RMSE statistics are only collected from the full 120-h free forecast period. An additional description will be added to the revised manuscript.

Lines 240-240: Might the relatively coarse 15-km model horizontal resolution also be a possible factor contributing to the model's failure to capture cyclogenesis for two cases? Even though (in lines 106-107) you mention the Li and Pu (2014) study's finding limited improvement in TC cyclogenesis prediction with increased model resolution, it appears plausible to this reviewer that since moist convection generally plays an important role in TC cyclogenesis, your model's inability to resolve convective updrafts at 15-km grid spacing could also hinder its ability to capture TC cyclogenesis for some cases.

We recognize that the relatively coarse horizontal resolution may be a contributing factor, a possibility that certainly warrants more detailed investigation. Accordingly, we will expand the discussion in the revised manuscript to acknowledge this potential influence.

Line 277: I think you meant to write "noted" after "It should be..", right? Please correct this.

Thank you for your suggestion. We will correct the sentence in the revised manuscript.

Line 334: "In contrast, none of the three DA experiments reproduced the observed increase in relative humidity…". Recommend replacing "observed" with "GFS-analyzed" or something similar since the GFS gridded analysis cannot be considered to be an observation.

Thank you for your suggestion. We will correct it in the revised manuscript.

Lines 334-335: "In contrast, none of the three DA experiments reproduced the … intensification of vorticity and its upward extension." It's worth noting that all three experiments do show transient upward development of weak cyclonic relative vorticity all the way to 100 hPa around t=0 h, although this feature only lasts for about 12 hours for all three experiments. Perhaps the more notable difference between the three experiments' vortex structures and the GFS is their shallower cyclonic relative vorticity field after t=12 h, continuing through the end of the evaluation period at t=48 h.

Thank you for this insightful observation. You are correct; we will clarify the text to acknowledge the transient upward extension of vorticity that occurs early in the simulation (around t = 0 h). We will revise the manuscript to explicitly differentiate this short-lived feature from the sustained intensification that is present in the GFS analysis but fails to persist in our DA experiments after t = 12 h. This distinction more accurately reflects the differences in the vortex structures.

Figure 15: Could you please mark the center location of developing TC Ian on the panels as forecast by the different experiments? At first I was confused looking at the strong TC located near the Bahamas in the t=1 h and t=24 h panels, until I realized that I was looking at Fiona. Also, it might be worth plotting the observed storm position on Figs. 15e-h since you contrast its motion with that of the GFS_q14kBestRO storm in lines 378-380. Interestingly, the GFS_init storm shown in Fig. 15d also appears to have a southward bias in its motion that brings it into a less favorable environment, but unlike the GFS_q14kBestRO storm, this storm can still intensify.

We will update the panels to mark the center location of the developing TC. We apologize for the confusion regarding the strong TC visible in the t = 1 h and t = 24 h panels; that storm is Hurricane Fiona. Additionally, we will add the observed best track positions for Hurricane Ian to panels (g-h), which highlights the slight southward bias in the GFS_14kBestRO experiment. Regarding your note on storm motion and intensification, Figure R5 now provides a comparison between the init_GFS and GFS_q14kBestRO experiments at t = 120 h, illustrating that the GFS_q14kBestRO storm has made landfall.

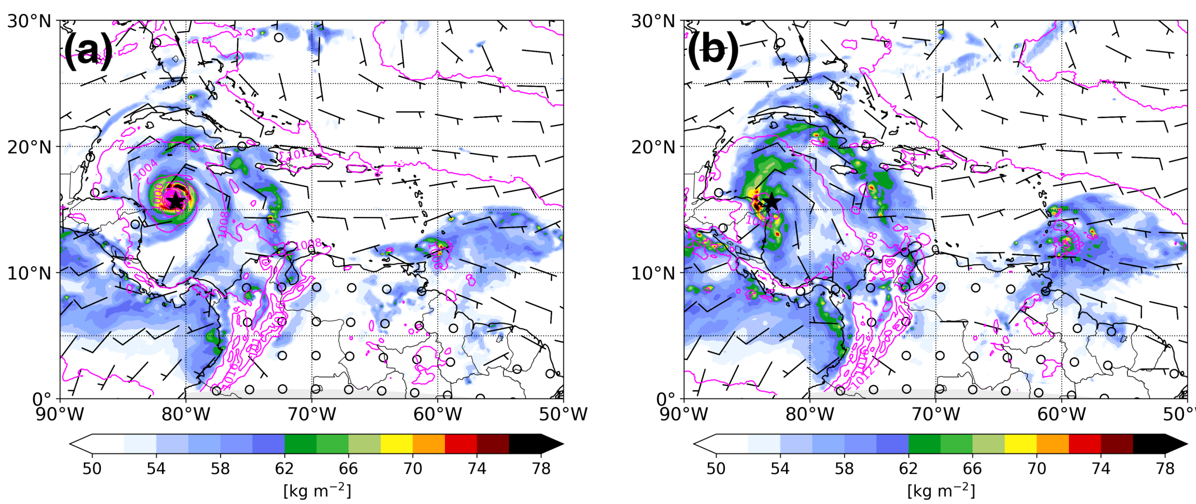

Figure R5: Snapshot of the mean sea level pressure (magenta contours; hPa), precipitable water (color-shaded; kg m^{-2}), and 10-m wind barb (black; kt) at forecast hour t = 120 h from (a) the init_GFS and (b) GFS_q14kBestRO experiments. The position of Hurricane Ian is indicated by a black star in each panel.

Line 414: ini_GFS —> init_GFS

Thank you! We will correct it in the revised manuscript.

Lines 414-417: By convective and non-convective precipitation, you are referring to precipitation generated by the cumulus physics and microphysics schemes, right? And not from a convective versus stratified partitioning algorithm, right? It would be good to mention this in the text or Figure 18 caption.

You are correct. The terms ‘convective’ and ‘non-convective’ precipitation refer to precipitation generated by the cumulus parameterization and microphysics schemes, respectively. We will update the caption of Figure 18 in the revised manuscript to explicitly state this distinction.