Response to Review 3 (RC3):

Comment: In this manuscript, the authors run a set of pseudo-global warming (PGW) simulations to investigate how 7 tropical cyclones (TCs) in the Pearl River Delta (PRD) region might change under warming perturbations from 16 different CMIP6 models (plus two additional "extreme" perturbations). They report results regarding storm pressure, wind, precipitation, and integrated kinetic energy. They find that storms generally increase in intensity across all metrics for 6 of the 7 storms. The application of climate storylines has become popular in recent years. The PGW framework is one way such storylines can be performed that targets regional scales. I agree that there is utility in such simulations, particularly around understanding how climate may change in these regions and communicating these to downstream individuals interested in climate data (e.g., planners, emergency managers, etc.).

Reply: We are pleased that the scientific contributions and implications of our study are well-recognized.

Comment: TCs and climate are important topics, but there are opportunities to deepen the analysis, better interpret the science, and make the findings more impactful. As is, I find the paper somewhat uninspiring. The actual analysis of the storms themselves is fairly shallow, with only basic qualitative comparative evaluation and not providing a deeper interpretation of the dynamical structure (e.g., how the changes in stability impact the storm's axisymmetric and asymmetric circulations). Even the evaluation of aspects such as the radius of maximum wind and precipitation distributions could be interesting. Conversely, the sample of 7 TCs doesn't feel large enough to draw any meaningful conclusions and the authors do not provide any real concrete reasons for their decision to eschew a larger sample size (28 storms). Of note, there is a lack of statistical tests to evaluate if any of the changes are robust. The general findings mainly serve as further confirmation of the numerous PGW/TC simulations published over the past 10 years or so.

Reply: We thank the Reviewer for encouraging us to enhance the quality of our study further.

Regarding the novelty of our study, we believe that combining the storyline approach with the pseudo-global warming approach enhances our understanding of uncertainties, particularly, in partitioning the uncertainties related to the physical aspects of global warming. Specifically, we have developed two additional storylines based on thermodynamic drivers to create scenarios with excessively *favorable and unfavorable conditions* for typhoon intensification extending beyond traditional GCM-driven simulations. The development of these two additional storylines allows us (1) to quantitatively understand the changes in associated uncertainties in simulated hydrometeorological variables (e.g., Figures 7 and 8) and (2) to enhance risk awareness by investigating deep uncertainties that go beyond the recognized limitations of GCM ensembles, particularly concerning the so-called “Gray Rhino” or “Black Swan” events in the context of weather and climate extremes. These statements are included to highlight the novelty of our study.

Regarding in-depth analysis, we closely follow the Review’s suggestions to examine how changes in stability impact the storm’s axisymmetric and asymmetric circulation. For this analysis, we exemplarily compare the vorticity fields from PGW-driven simulations with those from ERA5-driven simulations. Additionally, the temperature-precipitation relationship is investigated to better understand the Clausius-Clapeyron relation, particularly for typhoon cases under warming conditions.

The selection of the seven typhoons for the PGW investigation is based on the model performance evaluation results, which were assessed using objective metrics to reproduce the observed maximum wind speed and minimum sea level pressure. Since our study specifically focuses on warming-induced changes in simulated typhoon intensity, the selected typhoons encompass the full range from Category 1 to Category 4. For each typhoon, we have conducted 18 warming scenario simulations, and we believe that this sample size is sufficient to conclude each typhoon (shown in Figure 6). Of course, we are fully aware of the uncertainties in the simulations stemming from model physical schemes, configurations, synoptic situations, and warming scenarios. However, our preliminary simulations indicated that it is challenging to find a single model setup that performs well across all 28 typhoons. Therefore, our PGW analyses are restricted to the seven selected typhoons.

The above descriptions about the selection process for these seven typhoons are added into the revised version.

Regarding statistical tests, the non-parametric Mann-Whitney U-test is applied for quantitative significance testing, since the Gaussian distribution could not be assumed for all results in our study.

Comment: The paper reads a bit like using a tool (PGW simulations) to find a nail (TCs in the PRD). It currently is written in a way that reads very linearly and more like a technical document than a scientific paper. As is, I think the paper needs a fairly large overhaul before it can be considered for formal publication in a journal. The good news is I think this can be achieved with a better interpretation and deeper understanding of the data as opposed to additional model simulations. My major comments are described below and I think need to be critically addressed before any publication.
Reply: In this study, we employed a storyline approach to shed light on the warming induced changes in the simulated typhoon intensities. Our methodological proof-of-concept demonstrates the potential of combining the storyline approach with the pseudo-global warming approach, allowing us to frame risk in an event-oriented manner and eventually providing a physical basis for partitioning uncertainties.

We thank the Reviewer for positively providing us with the opportunity to enhance the quality of the manuscript.

Comment: Of note, the data availability requirements of HESS (https://www.hydrology-and-earth-system-sciences.net/policies/data_policy.html) do not appear to be satisfied by the statement "All data used to conduct this study will be made available upon request." The simulation data (or at least a subset of it able to recreate the results) should be uploaded to Zenodo or some other repository. Given the relatively small nature of the simulations (i.e., regional domain and short time windows), this shouldn't be an issue with some careful consideration of exactly what data should be archived.
Reply: We thank the Reviewer for introducing us to the Data Policy of HESS and the channels to archive the outputs of our study.

In this revision we adhere closely to the Data policy of HESS. Specifically, we upload our model setup (i.e., namelist files), visualization scripts, and key model outputs to the Zenodo repository to support open science.

Major comments:
Comment: There isn't an objective metric for choosing 7 of the 28 storms. The authors describe evaluations of intensity and track, but only superficially and rapidly jump from Fig. 4 (all TCs) to Fig. 5 (the 7 TC subset). The obvious concern here is -- if the other 21 TCs were poorly enough simulated to not be included in this analysis, is the version of WRF applied here really "fit for purpose"? That is, if it struggles to simulate the TCs, why do we believe the PGW signal is credible? Simply put, I think this decision needs a clearer rationale. If the other storms were excluded due to poor simulation quality, it would be helpful to provide a detailed analysis of these issues and discuss how they might affect the credibility of the remaining PGW simulations.
Reply: We utilize objective metrics, i.e., mean bias and typhoon category, to select seven out of the 28 storms for our study. The selection of these seven typhoons for the PGW investigation is based on the evaluation results of model performance in reproducing the observed maximum wind speed and minimum sea level pressure. Given that our study focuses on warming-induced changes in the simulated typhoon intensity, the selected typhoons represent the full range from Category 1 to Category 4. Of course, we are fully aware of the uncertainties in the simulations stemming from model physical schemes, configurations, synoptic situations, and warming scenarios. However, our preliminary simulations indicated that it is challenging to find a single model setup that performs well

across all 28 typhoons. Therefore, our PGW analyses are restricted to the seven selected typhoons, for which our WRF model setup demonstrates the best performance.

We believe that our derived PGW signals are credible, as they are based on the state-of-the-art Coupled Model Intercomparison Project CMIP6. To account for the uncertainties stemming from warming scenarios, we have conducted 18 warming scenario simulations for each typhoon, enabling us to draw our conclusion statistically (Figure 6).

In this revision additional PGW simulations for one selected typhoon (i.e., one with poorer performance) are conducted to further evaluate whether the performance of the WRF model affects the credibility of the remaining PGW simulations.

The above discussions about the selection of analyzed typhoons have been added into the revised "Results" section.

Comment: There is no statistical significance testing. This appears to be a fairly big oversight in my mind -- even basic tests like KS or a t-test could shed some light on how robust these changes are. The authors speculate about this occasionally in the manuscript (talking about the "majority" of members that increase in intensity, for example) but more quantitative analysis is needed.
Reply: We agree with the Reviewer about the need for statistical significance testing of our results.

In this revision the non-parametric Mann-Whitney U-test is used for quantitative significance testing, since the Gaussian distribution could not be assumed for all results in our study.

Comment: The IKE section is very underdeveloped (only 15 lines in the text). The evaluation is very superficial and, in my opinion, could be improved with fairly little effort. For example, Neoguri has a larger signal in the maximum wind versus the IKE. This would imply the structure of the TC is compensating (smaller?) for the IKE to change little even in the face of a large change in wind speed. Conversely, the IKE signal seems larger for Usagi, implying that the storm wind field is expanding (either in the inner core or the outer reaches). This comes back to my overarching concern that this paper reads very much as a shallowish description of model simulations without some deeper probing.
Reply: We thank the Reviewer for the guidance in enhancing the quality of our manuscript.

In this revision the IKE section is expanded in more detail. Following the Reviewer's suggestion, warming-induced changes in the structure of typhoons are investigated, such as Neoguri and Usagi, using metrics like vertical velocity and relative vorticity. Based on the findings from this comparative analysis, a similar approach is adopted to deepen our understanding of other typhoon PGW simulations derived from individual CMIP6 models.

Comment: In general, the paper could be better served by providing at least some spatial evaluation (e.g., 10.1038/s41586-018-0673-2). For example, is the precipitation maximum just increasing because all precipitation rates are increasing, or is the fundamental structure of the TC changing? In the storms that weaken in the PGW runs, does this appear to just be internal variability in storm intensity (see literature surrounding rapid intensification and weakening in TCs due to inner core processes) or is this fundamentally a response to the large-scale environment (perhaps an evaluation of metrics such as maximum potential intensity would be worthwhile)?
Reply: Thank you for sharing the references.

In the revised version, two new analyses are performed:

- A new analysis of changes in the spatial patterns of typhoon-induced precipitation under warming conditions is performed. This analysis helps us determine whether the increase in maximum precipitation is a systematic trend or is related to changes in the structure of the typhoons.
- A new analysis comparing intensified typhoon PGW runs with weakening typhoon PGW runs is performed using metrics, for example, maximum potential intensity. This comparison

provides insights into the physical mechanisms behind the differing responses (i.e., intensification versus weakening) of typhoons under warming conditions.

Minor comments:
Comment: Lines 106-107. Are we sure CMIP6 versus CMIP5 provides value-added for PGW runs?
Reply: We have rephrased this sentence as follows:
"In particular, our study emphasizes the use of the latest CMIP6 ensembles and the strategy of aligning these ensembles for future projections of tropical cyclones."

Comment: Fig. 2. While there is a sample of 16 GCMs. There really should be a discussion of model independence -- for example, there are multiple versions of CMCC, MPI, and NorESM2, with the only apparent difference being resolution. I would expect these sets of models to have very similar changes when compared to models with vastly different structural characteristics.
Reply: Yes, the inclusion of multiple versions of CMCC, MPI, and NorESM2 in our study facilitates comparisons of the simulations both within similar GCM frameworks and across different GCM frameworks.

In the revision we discuss our results in terms of model dependence and independence to address the uncertainties stemming from the GCMs. For the model dependence analysis, the simulations of CMCC, INM, MPI, and NorESM2 are selected. For the model independence analysis, the simulations of CanESM5 and CAMS-CSM1-0 are selected.

Comment: Line 229: 5 km grid spacing with cumulus convection remaining on? At these grid spacings, do researchers typically apply cumulus parameterizations or turn them off?
Reply: Yes, we turned on the Kain-Fritsch (KF) cumulus scheme for the 5-km grid spacing simulations.

We acknowledge that a grid spacing of 5 km is generally considered to be within the gray zone for simulating precipitation. This implies that the necessity for a cumulus scheme at this resolution can vary depending on the specific case. For example, Arnault et al. (2019) turned off cumulus parameterization in WRF to accurately reproduce a high-precipitation event in the upper Danube River basin in Europe. In contrast, Delfino et al. (2022) enabled the KF scheme while simulating Typhoon Haiyan over the North West Pacific using WRF.

In our study, we have conducted a series of preliminary simulations with the KF scheme both enabled and disabled. Our results indicate that the WRF model performs better when the KF scheme is enabled, which is in line with findings from sensitivity studies on the WRF model (Delfino et al., 2022; Sun et al., 2024).

The above discussion about the necessity of turning the KF scheme on at a 5 km grid spacing for simulating typhoons has been added into the revised "2.3 Methods" section.

Reference:
Arnault, J., Wei, J., Rummler, T., Fersch, B., Zhang, Z., Jung, G., et al., 2019: A joint soil-vegetation-atmospheric water tagging procedure with WRF-Hydro: Implementation and application to the case of precipitation partitioning in the upper Danube River basin. Water Resources Research, 55, 6217–6243. https://doi.org/10.1029/2019WR024780

Sun, Q., Olschewski, P., Wei, J., Tian, Z., Sun, L., Kunstmann, H., and Laux, P., 2024: Key ingredients in regional climate modelling for improving the representation of typhoon tracks and intensities, Hydrol. Earth Syst. Sci., 28, 761–780, https://doi.org/10.5194/hess-28-761-2024

Delfino, R. J., Bagtasa, G., Hodges, K., and Vidale, P. L., 2022: Sensitivity of simulating Typhoon Haiyan (2013) using WRF the role of cumulus convection, surface flux parameterizations, spectral nudging, and initial and boundary conditions, Nat. Hazards Earth Syst. Sci., 22, 3285–3307, https://doi.org/10.5194/nhess-22-3285-2022

Comment: Fig. 4. The authors should include a discussion of how the pressure-wind relationship simulated in models impacts these results (e.g., doi:10.6057/2018TCRR04.01)
Reply: Yes, the pressure-wind relationship in typhoons establishes a statistically meaningful link between the environmental conditions and the center of the typhoon, as well as the increase in the maximum surface wind around the storm (Bao et al., 2012).

In this revision the pressure-wind relationship in our typhoon simulations is investigated, and the derived results are added into the "Results" section.

Reference:
Bao, J.W., Gopalakrishnan, S.G., Michelson, S.A., Marks, F.D., Montgomery, M.T., 2012, Impact of physics representation in the HWRFX on simulated hurricane structure and pressure-wind relationships, Monthly Weather Review, 140, 3278-3299, https://doi.org/10.1175/MWR-D-11-00332.1

Comment: Fig. 6. I assume for the precipitation rates (c,d), the authors first find a single value at each model timestep (e.g., an array that is 1 x ntimes for each member) and then include those in the statistics. That is, there is no inclusion of spatial components in these boxplots?
Reply: We do incorporate spatial information when deriving the statistics shown in Figure 6. For instance, the maximum 1-hour precipitation rates (Figure 6c) represent the local maximum values surrounding the storm center within the 4° × 4° evaluation region.

The explanation of our analysis strategy is added in the "Data and Methods" section.

Comment: Figs. 7-8. It is very difficult to interpret these results outside of the thicker black, red, and blue lines. If the authors would like to provide more model-specific context, I would suggest either converting these lines to some form of shading or larger figures that could provide the ability to see some of the model lines (which currently are stacked very much on top of one another).
Reply: We appreciate the Reviewer's feedback.

We have improved the readability of Figures 7-8 as follows.

- Replace the individual lines with shaded bands to better visualize the uncertainties stemming from GCMs;
- Retain selected model lines within the shaded bands for the model independence and dependence analysis, such as the simulations from CanESM5 and CAMS-CSM1-0.

Comment: Lines 422. I am not sure what is particularly unique about Hato and Hagupi. Is it that the median value of the PGW runs falls in the 25-75% range of the ERA5?
Reply: Here, we are referring to the comparison between the colored circles (representing the warming scenario runs) to the black circle (representing the baseline run) in Figure 6c. The circles represent the maximum of the 1-hour precipitation extremes throughout the analyzed period. Figure 6c shows no clustering effects for Hato and Hagupit, as the colored circles (warming scenario runs) are widely and evenly distributed around the black circle (baseline run). This is what we meant by "no clear picture of a change in precipitation intensity".

Our detailed explanation about Figure 6 has been added into the revised Results for better understanding.