General comment

This manuscript investigates the sensitivity of stratospheric gravity waves generated by Typhoon Soudelor (2015) to WRF physical parameterizations, including microphysics, planetary boundary layer schemes, and the use of a cumulus parameterization in the outer domain. The study addresses an important problem. Tropical cyclones are strong sources of gravity waves, and the representation of convective forcing, wave propagation, and momentum transport remains a major uncertainty in numerical weather prediction and climate models. A process-oriented analysis of how model physics affects tropical-cyclone-generated gravity waves can therefore be valuable to the community.

However, in its current form, the manuscript remains too descriptive, lacks sufficient physical interpretation, and the experimental design is not fully justified in relation to the stated research question. I therefore recommend publication only after major revision.

Major issues

1. **Insufficient physical interpretation of the sensitivity results**

   The manuscript addresses an important question: how WRF physics choices affect tropical-cyclone-generated stratospheric gravity waves. However, the current analysis remains largely descriptive. The authors show that different microphysics, PBL, and cumulus configurations produce different cloud structures, latent-heating profiles, and gravity-wave responses, but the physical chain linking these differences is not yet demonstrated clearly enough. Without this mechanistic interpretation, the study mainly shows that the simulations differ, but does not fully explain why they differ.

2. **Ambiguity in the experimental design**

   The motivation for the experiment design needs clearer explanation. The study includes five PBL schemes, two microphysics schemes, and only one cumulus-parameterization configuration. Since the interpretation focuses mainly on differences among WSM6, Goddard, and WSM6–Grell-3, the role of the five PBL schemes is not fully clear. Are the PBL schemes intended as physically interpretable sensitivity experiments, or mainly as an ensemble sampling of uncertainty around the microphysics and cumulus choices? If the former, the authors should explain the physical pathways by which PBL formulation affects convection and gravity-wave generation. If the latter, this should be stated explicitly and individual PBL-scheme differences should be interpreted more cautiously. The cumulus-parameterization result also needs qualification. Because Grell-3 is tested only with WSM6 and not with Goddard, the conclusion that cumulus parameterization consistently weakens resolved convection and gravity-wave amplitudes should be restricted to the WSM6-based comparison unless an additional Goddard–Grell-3 experiment is added.

3. **Limited observational constraint on the gravity-wave response**

   The AIRS observations provide useful evidence that Soudelor generated stratospheric gravity waves, but they are used mainly as motivation rather than as a quantitative constraint on the WRF simulations. At minimum, the authors should compare model output at the times and locations of the AIRS overpasses and discuss whether the simulated wave packets have horizontal and vertical scales that would be visible to AIRS. Ideally, model temperature perturbations should be processed in a way that approximates AIRS sensitivity, or at least compared within the AIRS-relevant altitude and wavelength ranges. This would make the observational component more meaningful and would strengthen the claim that this is a well-observed case study.

Other comments:

The Introduction should include a clearer discussion of convective gravity-wave source spectra, transient diabatic heating, vertical harmonics of heating, and background-wind filtering. It should also explain more clearly why model-physics sensitivity of gravity waves is a distinct problem from model-physics sensitivity of tropical-cyclone track and intensity.

The introduction and methods would benefit from a more complete theoretical and methodological framing. Several claims are supported by citations, but in some places the references are too compressed or the relevant context is not sufficiently explained or cited. For example, line 178, the S-transform is introduced as a key diagnostic method for the first time, but the manuscript should cite both the relevant gravity-wave applications and, ideally, the original methodological reference here, while explaining why this method is appropriate for localized tropical-cyclone-generated waves. Similarly, lines 201-202, The AIRS sensitivity to middle- and upper-stratospheric gravity waves should also be cited more explicitly, including its altitude sensitivity and wavelength limitations.

Section 2.2: The description of AIRS observations could be more focused. General instrument details can be shortened, while details directly relevant to gravity-wave interpretation should be emphasized. The authors should also clarify whether the AIRS brightness-temperature perturbation method follows Hoffmann et al. (2013, 2014) exactly, or whether any modifications were made to the method. If the authors modified any part of the method, these modifications should be stated explicitly. At present, it is not fully clear whether the manuscript simply applies an established method or introduces methodological changes on top of previous work.

Figure 2: In panel (c), there appears to be a strong convective/cloud feature toward the northeastern part of the domain, but it does not seem to be associated with an equally clear wave-like perturbation in panel (d). Could the authors briefly clarify this? For example, is this related to AIRS viewing geometry or off-nadir sensitivity, the vertical/horizontal wavelength sensitivity of AIRS, background-wind filtering, timing between convection and wave propagation, or the possibility that this convective feature did not generate waves detectable by AIRS? This is not necessarily a problem, but a short explanation would help the reader interpret the cloud–wave relationship in Fig. 2.

Section 2.3: Please specify the WRF model time step. The manuscript reports the output frequency, but this is not the same as the integration time step. Please also clarify whether the nested domains are one-way or two-way interactive, whether Domain 02 was active for the full 72 h simulation, and whether any spin-up period was excluded from the analysis. Since the study focuses on convection, latent heating, and stratospheric gravity waves, early model adjustment of the vortex, hydrometeors, convection, and wave field could affect the results.

Line 125: Please provide a supporting reference for the 10 km threshold.

Section 2.4.2: The latent-heating-rate diagnostic needs clarification. The manuscript states that LHR is calculated from temporal tendencies of hydrometeor mixing ratios using $\Delta t = 900$ s, corresponding to the model output frequency. Please clarify whether the Δq terms are true microphysical tendency terms from WRF or finite differences of archived hydrometeor fields. If they are finite differences over 15-minute output intervals, the diagnostic may include advection, sedimentation, turbulent mixing, numerical diffusion, and hydrometeor transport, and therefore may not represent pure latent heating from phase changes. In that case, the term LHR should be used more cautiously and the limitation should be discussed.

Lines 190–195: Please specify the KDE bandwidth used in the analysis. Since KDE smoothing can affect the apparent shape of the distributions, including peaks, tails, and possible multimodality, the bandwidth choice should be justified, or the authors should show that the conclusions are not sensitive to reasonable bandwidth changes.

Section 3.2: The comparison of simulated track and intensity with best-track data is useful, but the discussion could be strengthened by placing the results more clearly within the broader WRF tropical-cyclone sensitivity literature. For example, are the track divergence near landfall, delayed weakening after land interaction, and stronger intensity in Goddard simulations consistent with previous WRF studies of tropical cyclones?

Section 3.3: The OLR comparison with FengYun-2G observations is useful, but the interpretation should acknowledge that model–observation discrepancies may not arise only from physics parameterizations. Since the simulations are initialized and forced by ERA5 rather than constrained by FengYun-2G cloud observations, some discrepancies may reflect initial-condition and boundary-condition errors. The authors can still use the satellite comparison to evaluate realism, but they should be cautious when attributing absolute OLR differences solely to physics choices.

Lines 318–320: The statement that heating is more intermittent and localized during intensification, and more spatially distributed and temporally smoother at peak intensity, is plausible but not fully demonstrated by the current analysis. Since the LHR profiles are based on grid points exceeding the 95th percentile, the method emphasizes the strongest heating cores and does not by itself quantify spatial distribution or temporal smoothness. Please either provide additional diagnostics, such as LHR maps, area fractions above thresholds, PDFs, time series, or intermittency metrics, or soften this interpretation.

Section 3.5: The standard deviation of vertical velocity is a useful proxy for gravity-wave activity, especially in the stratosphere, but it is not sufficient to assess the dynamical impact of the waves. The manuscript refers to terms like momentum transport and vertical coupling, but does not quantify momentum flux, wave energy, or wave drag.

Figure 10: The first several hours of the vertical-velocity-variability time series show very low values before the main variability develops. This may reflect model spin-up or convective

adjustment. Please clarify whether early adjustment affects the interpretation of the subsequent gravity-wave time series.

Lines 361–362 and Fig. 10: The MYJ scheme produces larger vertical-velocity variability, especially when combined with Goddard microphysics. However, this large PBL-related spread should be interpreted with caution. PBL schemes can alter low-level winds, inflow, land–sea interaction, and interaction with Taiwan's topography. As the storm approaches Taiwan, orographic gravity waves may also contribute to stratospheric vertical-velocity variability if propagation conditions allow.

The relationship between the LHR diagnostics and the gravity-wave response should be made more quantitative. For example, Goddard–BouLac appears to have relatively strong heating characteristics, whereas Goddard–MYJ produces the largest vertical-velocity variability. This suggests that LHR magnitude alone may not explain the simulated gravity-wave response. A clearer analysis linking heating structure, source spectrum, propagation conditions, and wave amplitude would strengthen the causal interpretation.

Line 396: The statement that differences are mostly attributable to diabatic-heating structure and convective forcing seems too strong based only on the similarity of domain-mean wind and temperature profiles. Propagation and filtering depend on wind, stability, and critical levels, and domain-mean profiles may hide important local differences. Please either provide additional evidence for this attribution or soften the statement.

Lines 401–404: The amplitude threshold of 0.3 $m\ s^{-1}$ needs stronger justification. The appendix shows threshold sensitivity in terms of spatial coverage, but the main spectral conclusions may still depend on the threshold. Large vertical-velocity amplitude does not necessarily imply the largest contribution to vertical momentum or energy transport. The authors should discuss how the threshold affects the interpretation, or include diagnostics that consider a broader wave-amplitude spectrum.

Lines 486–495, 503-507: The interpretation would benefit from additional analysis of background-wind filtering and gravity-wave propagation conditions. Heating depth and altitude are important for the generated source spectrum, but the waves that reach the analysed stratospheric levels also depend on background wind, stability, and critical-level filtering. The similarity of domain-mean wind and temperature profiles in Fig. 13 is not sufficient by itself to rule out propagation effects.

Conclusions: The conclusions should be made more cautious. The study supports strong sensitivity of simulated gravity-wave activity to model physics for this case, but it does not yet fully establish general behaviour across tropical cyclones or quantify the impact on the mean flow.

Textual and figure comments:

Abstract: The statement that the differences “demonstrate” control by diabatic forcing is too strong. “Suggest” or “indicate” would be more appropriate unless additional source–propagation diagnostics are added.

Please use the abbreviation “TC” consistently after defining tropical cyclone. Avoid introducing the abbreviation and then repeatedly switching back to the full term where the abbreviation would be clearer.

Figures 2 and 3a: It would be helpful to add time labels to the storm track, at least at selected intervals, so that the reader can more easily relate the location of the convective core to the observed gravity-wave perturbations and to the simulated storm evolution.

Figures 1–6: Please make the longitude labels consistent across the figures. Some plots label longitudes with “E” notation, for example Figs. 1 and 3–6, while Fig. 2 appears to omit this convention. Please use the same longitude/latitude labelling style throughout the manuscript.

Figure 8: Please include the simulation time in the panel titles, not only in the caption. This would make the figure easier to read.

Please use a consistent colour and line-style convention for the same simulations throughout the manuscript. At present, the convention appears to differ among Figs. 3–7, 8–9, 10, and 13, which makes it difficult to follow individual configurations across diagnostics.

Figures 11–12: Please overlay the storm-centre location on the maps.

Figure 13: Please make the ERA5 line style consistent between the plot, caption, and legend. The caption refers to ERA5 as a black line, but it appears to be a black dashed line; also ERA5 should be included explicitly in the legend. Please also explain what the highlighted 95 hPa level represents. In addition, it would be useful to mark the two analysed gravity-wave levels, approximately 21 km and 39 km, on these profiles so that the reader can relate the background wind and temperature structure directly to the levels used in the gravity-wave analysis.

References:

- Hoffmann, L., Xue, X., and Alexander, M. J.: A global view of stratospheric gravity wave hotspots located with Atmospheric Infrared Sounder observations, Journal of Geophysical Research: Atmospheres, 118, 416–434, https://doi.org/10.1029/2012JD018658, 2013.
- Hoffmann, L., Alexander, M. J., Clerbaux, C., Grimsdell, A. W., Meyer, C. I., Rößler, T., and Tournier, B.: Intercomparison of stratospheric gravity wave observations with AIRS and IASI, Atmospheric Measurement Techniques, 7, 4517–4537, https://doi.org/10.5194/amt-7-4517-2014, 2014.