Reply to Reviewer #2

We sincerely thank the reviewer for the careful evaluation of our manuscript and for the constructive comments and suggestions, which have helped us to improve the quality, clarity, and scientific presentation of the paper. We have carefully considered all comments and revised the manuscript accordingly. Please find our detailed replies to the reviewer's comments below, with our responses shown in blue.

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.

We thank the reviewer for this careful and constructive overall assessment of our manuscript. We agree that the original version was too descriptive in several places and did not sufficiently justify the experimental design in relation to the stated scientific objective.

In response, we have substantially revised the manuscript in three main respects. First, we have strengthened the physical interpretation throughout the Introduction, Methods, Results, and Discussion. The revised text now explains more explicitly how differences in microphysics, PBL formulation, and outer-domain cumulus parameterization can alter convective organization, the magnitude and vertical structure of diabatic heating, and thereby the generation and subsequent stratospheric characteristics of tropical cyclone-generated gravity waves. We also discuss more clearly the role of background winds, stability, and possible critical-level filtering during upward propagation.

Second, we have expanded and clarified the experimental design. Thompson microphysics has been added as a third microphysics option, and the Grell–3 cumulus parameterization is now tested in combination with all three microphysics schemes. We have also clarified that the PBL experiments are used to sample uncertainty in the low-level thermodynamic and momentum supply to the storm, rather than being interpreted as parameterizations that directly control gravity wave generation. The revised design therefore addresses three complementary physical pathways: microphysical control of latent heating structure, *outer-domain cumulus control of the environment supplied to the convection-permitting nest, and PBL control of the low-level forcing of deep convection.*

Third, we have strengthened the gravity wave diagnostics. In addition to the vertical velocity variability used in the original manuscript, we now document the localized S-transform analysis with a representative example showing the diagnosed wave amplitude and horizontal and vertical wavelengths. We have also added a direct perturbation-based estimate of gravity wave momentum flux from the simulated wind fields. These additions allow us to assess both the localized spectral characteristics of the waves and the vertical flux of horizontal momentum.

At the same time, we have moderated the causal wording throughout the manuscript. Where the available diagnostics do not permit a unique separation of source differences from propagation and filtering effects, we now state this explicitly and present the proposed mechanisms as physically consistent interpretations rather than definitive process-level proof.

We believe that these revisions have substantially improved the physical basis, methodological clarity, and overall balance of the study and have addressed the main concerns raised in the general comment.

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.

We thank the reviewer for this important comment. We agree that the previous version described the sensitivity of the simulations more clearly than it explained the physical mechanisms linking the different model physics configurations to the resulting stratospheric gravity wave response.

We have therefore substantially expanded the physical interpretation throughout the manuscript. In the model-description section, we now summarize the theoretical and structural differences among the selected microphysics, PBL, and cumulus parameterizations that are relevant to tropical cyclone convection and gravity wave generation. In particular, we explain how microphysical differences in hydrometeor production, phase conversion, and sedimentation can affect the magnitude and vertical distribution of diabatic heating; how PBL closures can alter turbulent mixing, inflow, moisture convergence, and convective organization; and how the outer-domain Grell–3 cumulus scheme can indirectly modify the environment supplied to the convection-permitting inner domain.

The Results section has also been revised to make the physical chain more explicit. The differences in simulated cloud-top structure are now related to changes in the strength, localization, and vertical placement of latent heating. These heating differences are subsequently interpreted in terms of their projection onto the available gravity wave modes, providing a physically *consistent explanation for differences in gravity wave amplitude and spectral characteristics. The revised text also explains that the PBL influence is indirect and depends on interactions with microphysics, storm evolution, and the environmental flow.*

Finally, we have expanded the Discussion to synthesize these mechanisms into three principal pathways: microphysical control of the vertical structure of latent heating, indirect modification of resolved convection by outer-domain cumulus parameterization, and PBL control of the low-level thermodynamic and momentum supply to the storm core.

We have also added two complementary diagnostics that strengthen the connection between the simulated wave structures and their dynamical relevance. A new S-transform example figure explicitly shows the locally dominant wave amplitude and the associated horizontal and vertical wavelengths for a representative simulation. In addition, a new perturbation-based momentum flux analysis quantifies the vertical flux of horizontal momentum using $\rho u'w'$ *and* $\rho v'w'$*. The resulting domain-mean and radial GWMF comparisons demonstrate that the model-physics sensitivity is also evident in a momentum-transport quantity, although the relative contrasts are not identical to those found for vertical velocity variability or positive-definite wave-amplitude measures.*

Where the available diagnostics do not permit a unique process-level attribution, we now state this limitation explicitly and present the proposed mechanisms as physically consistent interpretations rather than definitive proof of individual parameterized processes.

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.

We agree with the reviewer that the original experimental design was imbalanced, with substantially greater sampling of PBL schemes than of microphysics and cumulus parameterization. We have therefore expanded the experiment set in two ways.

First, Thompson microphysics has been added as a third microphysics option, in addition to WSM6 and Goddard. Second, the Grell–3 cumulus parameterization is now tested in combination with all three microphysics schemes. These additional simulations allow us to assess whether the response to outer-domain cumulus parameterization is robust across different microphysical representations rather than being inferred only from the WSM6-based comparison.

We have also clarified the role of the PBL experiments. The PBL schemes are used to sample uncertainty in the low-level thermodynamic and momentum supply to the storm rather than being interpreted as parameterizations that directly control gravity wave generation. Through their effects on turbulent mixing, surface inflow, moisture convergence, and convective organization, the PBL schemes can indirectly modify the structure and intermittency of diabatic heating and therefore the gravity wave source.

The revised experimental design thus addresses three complementary physical pathways: microphysical control of condensate and latent heating structure, outer-domain cumulus control of the environment supplied to the convection-permitting inner domain, and PBL control of the low-level forcing of deep convection. We have revised the Methods and Discussion accordingly and now interpret individual PBL-scheme differences more cautiously.

The new results further show that the effect of Grell–3 is not uniform across the three microphysics schemes. We have therefore removed the previous claim that cumulus parameterization consistently weakens resolved convection and gravity wave amplitudes and now describe the response as configuration dependent.

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 thesimulated 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.

We agree with the reviewer that the observational role of the AIRS data needed to be clarified. The original intention of including AIRS was to provide observational context for the selected case and to demonstrate that stratospheric gravity wave activity associated with Typhoon Soudelor was indeed detected by an independent satellite instrument. It was not intended as a quantitative validation of the complete simulated gravity wave spectrum.

In response to the reviewer's comment, we examined the requirements for a more quantitative model–observation comparison. Such a comparison would require the WRF temperature perturbations to be processed with the AIRS observational filter, including the broad vertical weighting functions, horizontal sampling, viewing geometry, and brightness temperature response. A direct comparison between unfiltered model fields and AIRS brightness temperature perturbations would otherwise not compare equivalent quantities.

This distinction is particularly important for the present case. AIRS is preferentially sensitive to gravity waves with relatively long vertical wavelengths, approximately $\lambda_z \gtrsim 15$–20 *km, because waves with shorter vertical wavelengths are strongly attenuated by the broad vertical weighting functions of the nadir-viewing 4.3 μm channels. In contrast, our S-transform analysis shows that the locally dominant vertical wavelengths in the WRF simulations are frequently shorter than this AIRS-sensitive* *range. AIRS therefore observes only a filtered subset of the simulated gravity wave spectrum, and this subset is not generally the dominant component diagnosed in the simulations.*

For this reason, we concluded that a quantitative comparison based only on wave amplitudes or dominant wavelengths would have limited physical meaning unless an appropriate AIRS observational operator, including radiative transfer and spatial sampling, were applied. We have therefore not presented the AIRS observations as a quantitative validation of the WRF wave amplitudes or spectra.

Instead, we have revised Sects. 2.2 and 3.1 to state more explicitly that the AIRS observations are included as observational context. They provide independent observational evidence for stratospheric gravity wave activity during the Soudelor case and show wave-like perturbation structures in the vicinity of the storm, while the model-sensitivity analysis itself is based on the internally consistent comparison among the WRF simulations. We now also discuss explicitly the AIRS altitude and wavelength sensitivity and caution that the absence or weakness of an AIRS signal does not imply the absence of gravity waves in the simulations or in the atmosphere.

We agree that a future quantitative comparison using an AIRS forward model would be valuable, particularly for isolating the subset of simulated waves that would be observable by AIRS. However, such an analysis constitutes a separate instrument-simulation study and is beyond the scope of the present model-physics sensitivity investigation.

Other Comments

1. 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.

   We thank the reviewer for this suggestion. We have expanded the Introduction to provide a clearer theoretical framework for convectively generated gravity waves. The revised text now explains that transient diabatic heating and convective motions generate a broad source spectrum rather than a single characteristic wave. We also discuss how the vertical heating profile can be represented by vertical harmonics and how their projection onto the natural vertical modes of the atmosphere influences the preferentially excited wave components. In addition, we now describe how the background wind modifies the intrinsic frequency and vertical wavelength of the waves and selectively filters the source spectrum during upward propagation.

   We have also clarified why gravity wave sensitivity to model physics is distinct from conventional sensitivity of tropical cyclone track and intensity. Track and intensity are bulk storm measures, whereas the gravity wave source depends on finer-scale characteristics of convection, including the depth, altitude, organization, and temporal variability of diabatic heating. These characteristics may vary substantially among physics configurations even when the simulated storm tracks and intensities are similar. Relevant references on convective gravity wave generation, source spectra, vertical heating structure, and propagation filtering have been added.

2. 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.

We thank the reviewer for this helpful suggestion. We have strengthened both the methodological description of the S-transform and the discussion of AIRS gravity wave sensitivity.

In Sect. 2.4.3, we now cite the original methodological reference of Stockwell et al. (1996), in addition to previous gravity wave applications. We have also expanded the explanation of why the S-transform is appropriate for the present analysis. In particular, its spatially localized and scale-dependent representation allows the amplitude and dominant scales of individual wave packets to be identified, making it well suited to gravity waves generated by transient and spatially localized tropical cyclone convection.

In Sects. 2.2 and 3.1, we now cite Hoffmann et al. (2013, 2014) directly in connection with the altitude and wavelength sensitivity of the AIRS 4.3 μm channel combination. The revised text specifies that the weighting functions peak at approximately 30–40 km and that the nadir-viewing measurements are preferentially sensitive to waves with long vertical wavelengths and horizontal wavelengths within the approximate AIRS detection range. We also clarify that waves outside these ranges may be strongly attenuated or remain undetectable in the brightness temperature perturbations.

3. 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.

We thank the reviewer for this helpful comment. We agree that the original description included more general instrument information than was necessary while not stating sufficiently clearly how the brightness temperature perturbations were calculated. We have shortened the general description of AIRS and expanded the discussion of the aspects directly relevant to gravity wave interpretation, including the altitude range of the 4.3 μm channel combination, the approximate horizontal- and vertical wavelength sensitivity, and the limitations arising from the nadir viewing geometry.

We have also clarified that the brightness temperature perturbation processing follows Hoffmann et al. (2013, 2014) without modification. In particular, no changes were made to the selection and averaging of the 42 channels or to *the background-removal and perturbation-calculation procedure. The revised Section 2.2 now states this explicitly and briefly summarizes the relevant processing steps.*

4. 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.

   We thank the reviewer for pointing out the lack of a clear wave-like counterpart to the northeastern convective feature in Fig. 2c. We have added a short clarification to the manuscript. A one-to-one correspondence between individual convective systems and AIRS-detectable gravity waves is not necessarily expected, because different sources may generate different gravity wave spectra. The waves associated with this feature may have had horizontal or vertical wavelengths outside the AIRS sensitivity range. In particular, under weak stratospheric background winds, as frequently encountered at low tropical latitudes, the resulting vertical wavelengths may be too short to be resolved effectively by the broad vertical weighting functions of the AIRS 4.3 μm channels. Their brightness temperature amplitudes would consequently be strongly attenuated. Background-wind filtering may further reduce the part of the generated spectrum that reaches the AIRS-sensitive altitude range. We therefore do not interpret the absence of a clear perturbation pattern as evidence that this convective feature generated no gravity waves.

5. 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.

   We thank the reviewer for requesting these clarifications. We have revised Sect. 2.3 to state explicitly that the model integration time step is 18 s. We now also specify that Domain 02 is a moving nest with two-way feedback to Domain 01 and that it remains active throughout the full 72-h simulation.

   We agree that the initial adjustment of the vortex, hydrometeor fields, convection, and gravity wave field may affect the early simulation period. The first 14 h are therefore treated as model spin-up and are excluded from the main statistical interpretation. We have added this information to the Methods and have adjusted the relevant time-series figures and discussion so that the early adjustment period is clearly identified.

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

We thank the reviewer for pointing out that the 10-km threshold required supporting documentation. We have added the relevant WRF reference and revised the sentence to clarify that mixed-phase microphysics schemes are recommended for convection-permitting and convection-resolving grid spacings of approximately 10 km or finer. We also avoid presenting 10 km as a strict physical cutoff, since the appropriate treatment of convection and microphysics depends on the model configuration and application.

7. 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 Δt = 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.

We thank the reviewer for raising this important point. We confirm that the Δq terms are calculated as finite differences of archived hydrometeor mixing-ratio fields at the 15-min output interval rather than from the native WRF microphysical tendency terms.

We agree that this diagnostic may therefore include contributions from advection, sedimentation, turbulent transport, numerical diffusion, and other processes in addition to phase changes. It should consequently not be interpreted as a pure microphysical latent heating tendency.

We have revised Sect. 2.4.2 to describe the quantity more cautiously as an approximate latent heating or diabatic heating proxy. Because the same output interval and calculation procedure are applied to all simulations, the diagnostic remains useful for internally consistent inter-simulation comparisons, but we no longer attribute the diagnosed differences uniquely to microphysical phase changes.

8. 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.

We agree that the KDE bandwidth should be stated explicitly because the degree of smoothing can affect the apparent shape of the distributions. We have therefore added the bandwidth values to the Methods and the relevant figure captions.

The bandwidth was selected empirically to suppress small-scale sampling noise while retaining the principal peaks, tails, and differences among the simulations. We acknowledge in the revised manuscript that the detailed shape of individual KDE curves may depend on this choice and therefore avoid overinterpreting minor peaks or small-scale distribution features.

9. 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?

We thank the reviewer for this suggestion. We have expanded Sect. 3.2 to place the simulated track and intensity differences in the context of previous WRF tropical cyclone sensitivity studies.

Previous studies have shown that TC track is often less sensitive to physics choices than storm intensity and inner-core structure, although differences can increase near landfall when the simulated circulation interacts with coastal terrain, land-surface contrasts, and changes in environmental steering. Delayed weakening after land interaction has also been reported in simulations that maintain excessive inner-core convection or surface enthalpy supply. Likewise, the stronger intensities produced by some microphysics configurations are consistent with previous findings that hydrometeor conversion, condensate loading, latent heating distribution, and precipitation efficiency can substantially modify the simulated storm intensity.

We have added relevant references from the WRF tropical cyclone sensitivity literature and now make clear that the behavior found here is broadly consistent with earlier studies, while also noting that the precise response is case and configuration dependent.

10. 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.

We agree with the reviewer that discrepancies between simulated and observed OLR cannot be attributed solely to the selected physics parameterizations. The WRF simulations are initialized and laterally forced by ERA5 and are not constrained by the FengYun-2G cloud observations. Differences in the initial storm structure, environmental moisture, lateral-boundary forcing, and storm position may therefore also contribute to the model–observation differences.

We have revised Sect. 3.3 to make this limitation explicit. The satellite OLR comparison is now used to assess the overall realism and spread of the simulated convective cloud fields, rather than to assign the absolute model–observation discrepancies uniquely to individual physics schemes.

11. 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.

We agree that the previous interpretation was not fully supported by the percentile-based vertical profile diagnostic. Because the analysis includes only grid points above the 95th percentile, it characterizes the strongest heating cores but does not directly quantify their spatial coverage, intermittency, or temporal smoothness.

We have therefore removed the statements that heating is more localized and intermittent during intensification and more spatially distributed and temporally smoother at peak intensity. The revised text is restricted to the differences that are directly supported by the diagnosed vertical heating profiles, including their magnitude, peak altitude, and vertical extent.

12. 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.

We thank the reviewer for this comment and agree that vertical velocity variability alone is not sufficient to assess the dynamical importance of the simulated gravity waves. We have therefore added a dedicated subsection that introduces a direct perturbation-based estimate of gravity wave momentum flux from the WRF wind fields.

A large-scale horizontal background is removed from the horizontal and vertical winds using an FFT-based horizontal filter, yielding the perturbation fields u', v', and w'. The local momentum flux components are calculated as $\rho u'w'$ and $\rho v'w'$, and the principal comparison quantity is the magnitude of the horizontally averaged flux vector,

$$|\mathrm{GWMF}| = \sqrt{\langle \rho u'w' \rangle^2 + \langle \rho v'w' \rangle^2}.$$

The revised manuscript includes a representative map and vertical cross-section of the perturbation and GWMF fields. We also compare the horizontally averaged net GWMF measure among all available physics configurations and examine its radial structure relative to the simulated TC center. The calculation uses all valid grid points of the inner-domain analysis region and does not apply an additional geographical mask or vertical velocity-amplitude threshold.

These additions allow us to move beyond a purely proxy-based discussion of gravity wave activity. The S-transform is retained for the complementary purpose of identifying the locally dominant wave amplitudes and wavelengths, whereas the direct perturbation-based calculation is used for the momentum flux diagnosis.

We emphasize that the new diagnostic quantifies momentum flux rather than momentum deposition. It does not directly diagnose wave breaking, the vertical divergence of momentum flux, or gravity wave drag. We have revised the manuscript accordingly and no longer imply that the present analysis quantifies the resulting mean-flow forcing.

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

We agree that the very small vertical velocity variability during the first several hours primarily reflects model spin-up and convective adjustment. To prevent this early adjustment period from affecting the interpretation, the first 14 h are excluded from the main analysis.

We have revised Fig. 10 so that the displayed and interpreted time series begin after the spin-up period. Where the complete simulation period is shown for context, the excluded interval is clearly marked. The discussion now focuses only on the subsequent evolution of the gravity wave field.

14. 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.

We agree that the particularly large vertical velocity variability in the MYJ configurations should be interpreted cautiously. PBL schemes can modify low-level winds, surface inflow, moisture convergence, land–sea interaction, and the structure of the circulation approaching Taiwan. These changes may affect both the convective gravity wave source and the interaction of the storm circulation with the island topography.

We have therefore added a discussion noting that orographically generated gravity waves may contribute to the diagnosed stratospheric variability near Taiwan if the local propagation conditions are favorable. The present diagnostics do not uniquely separate convective and orographic wave sources. We consequently no longer interpret the enhanced MYJ variability as evidence of a purely convective PBL effect, but as a configuration-dependent response that may include interactions among PBL physics, storm evolution, topography, and background propagation conditions.

15. 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.

We agree that the original manuscript did not make the relationship between the diagnosed heating characteristics and the gravity wave response sufficiently quantitative. The reviewer is also correct that the magnitude of the latent heating proxy alone cannot explain the ordering of the simulated gravity wave responses.

We have therefore strengthened the revised manuscript in two ways. First, a new S-transform example figure explicitly documents the locally dominant wave amplitude and the corresponding horizontal and vertical wavelengths retrieved from the simulated stratospheric wave field. This makes the spectral characterization more transparent and illustrates that the analysis includes localized wave-scale information in addition to bulk vertical velocity variability.

Second, we have added a direct perturbation-based estimate of gravity wave momentum flux from $\rho u'w'$ and $\rho v'w'$. The resulting horizontally averaged net GWMF measure and its radial structure provide an additional quantitative description of the wave response that is dynamically more direct than the standard deviation of vertical velocity.

The new results also support the reviewer's observation that latent heating magnitude alone does not determine the gravity wave response. For example, the relative ordering of the physics configurations in the direct GWMF diagnostic is not identical to their ordering in the latent heating or vertical velocity diagnostics. We now interpret this as evidence that the wave response depends on the spatial and vertical structure, localization, and temporal variability of the convective forcing, as well as on the phase relationships among the wave perturbations and on subsequent propagation and filtering.

A complete quantitative closure linking the full convective source spectrum, the intervening background-wind and stability structure, and the spectrum at each stratospheric altitude would require a direction-resolved multi-level frequency–wavenumber analysis. Such an analysis remains beyond the scope of the present study. We have therefore moderated the causal wording throughout the manuscript and now describe latent heating as an important contributing factor rather than the sole control of the simulated gravity wave response.

16. 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.

We agree that the previous attribution was too strong. Similar domain-mean wind and temperature profiles do not exclude local differences in wind, stability, intrinsic phase speed, or critical-level filtering. We have therefore softened the corresponding statement and no longer attribute the simulated wave differences primarily or exclusively to diabatic forcing.

The new S-transform and momentum flux diagnostics provide a more quantitative description of the resulting stratospheric wave field, but they do not by themselves uniquely separate source differences from propagation effects. The revised manuscript now describes differences in convective and diabatic forcing as important contributors while explicitly acknowledging possible differences in propagation and filtering.

17. 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.

We agree that the fixed threshold of $0.3\ m\ s^{-1}$ was not sufficiently well justified and that the main physical interpretation should not depend on a single cutoff value. We have therefore revised the analysis strategy.

The new S-transform example figure shows the continuous diagnosed fields rather than only a thresholded wave subset. For the statistical S-transform analysis, the diagnosed wave properties are weighted by the local amplitude, which gives stronger waves greater influence without excluding weaker components entirely.

The new direct GWMF analysis is likewise evaluated over all valid grid points of the filtered perturbation fields and does not use a manually selected vertical velocity-amplitude threshold. The revised central conclusions therefore no longer depend on the previous $0.3\ m\ s^{-1}$ cutoff.

The previous threshold analysis has been removed.

18. 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.

We agree that the similarity of domain-mean profiles is not sufficient to rule out propagation effects. We have revised the Discussion accordingly and now state explicitly that local background-wind structure, stability, intrinsic phase speed, and critical-level filtering may contribute to the differences among the simulated stratospheric wave fields.

The newly added S-transform and GWMF diagnostics strengthen the characterization of the waves at the analysed altitudes, but they do not constitute a complete source-to-stratosphere propagation analysis. We therefore avoid interpreting the diagnosed differences as evidence of diabatic forcing control alone.

19. 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.

We agree and have revised the Conclusions to make the scope and limitations more explicit. The revised text now states that the results show substantial model-physics sensitivity for the single case of Typhoon Soudelor, rather than establishing universal behavior across tropical cyclones.

We have also distinguished more clearly between gravity wave activity, momentum flux, and momentum deposition. The newly added perturbation-based diagnostic quantifies differences in GWMF among the simulations, but the study does not diagnose wave breaking, flux divergence, gravity wave drag, or the resulting mean-flow response. Claims concerning the broader dynamical impact of the waves have therefore been moderated accordingly.

Textual and figure comments:

1. 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.

   We agree and have replaced the previous statement that the differences "demonstrate" control by diabatic forcing. The revised Abstract now states that the results suggest that differences in convective and diabatic forcing are important contributors to the simulated gravity wave sensitivity, while also acknowledging possible differences in propagation and filtering.

2. 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.

   We agree and have revised the manuscript to use the abbreviation "TC" consistently after its first definition. The full term is retained only where needed for readability or at the beginning of a section.

3. 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.

   We agree and have added time labels at selected intervals along the storm track in Figs. 2 and 3a. This makes it easier to relate the observed and simulated wave structures to the position and evolution of the TC.

4. 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.

   We agree and have revised Figs. 1–6 to use a consistent longitude and latitude labeling convention. Longitude labels now include the "E" designation throughout, including the AIRS panels in Fig. 2.

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

   We agree and have added the simulation time directly to the panel titles in Fig. 8, in addition to retaining the information in the caption.

6. 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.

We agree and have standardized the colour and line-style conventions used for the physics configurations throughout the manuscript. The same simulation is now represented consistently across Figs. 3–13 wherever possible, and the captions and legends have been updated accordingly.

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

We agree and have overlaid the simulated TC-center location on the maps in Figs. 11 and 12.

8. 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.

We agree and have revised Fig. 13 accordingly. ERA5 is now shown consistently as a black dashed line in the plot, caption, and legend.

We have also clarified that the highlighted 95-hPa level represents the mean tropopause pressure during the analysed period. The gravity wave analysis level near 21 km is now marked on the profiles. The higher analysis level near 39 km lies outside the displayed vertical range of this figure and is therefore identified in the caption rather than marked directly on the profiles.

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.

Stockwell, R., Mansinha, L., and Lowe, R.: Localization of the complex spectrum: the S transform, IEEE Transactions on Signal Processing, 44, 998–1001, https://doi.org/10.1109/78.492555, 1996.