Reviewer #1:

We sincerely thank the reviewer for the careful reading of our manuscript and for the constructive comments. We are grateful for the positive assessment that the diffusion-based filter achieves the intended scale-selective capability. We have revised the manuscript in several aspects. First, we have clarified the main goal of the study in the Introduction, emphasizing the development and evaluation of an online scale-selective nudging scheme for MPAS-A on an unstructured variable-resolution mesh. Second, we have reorganized Sect. 2.2 to better describe the diffusion-based filtering framework following Grooms et al. (2021), and corrected the notation and equations related to the discrete Laplacian operator. Our responses to the reviewers' comments are inserted after each comment below in blue text and preceded by the word "Response". We believe these revisions have improved the clarity of the manuscript.

In the introduction, I would suggest that the author use a few words explicitly introducing the goal of this study. The author should state it more clearly so the reader can follow more easily.

Response: We thank the reviewer for this helpful suggestion. We agree that the goal of the study should be stated more explicitly in the Introduction. The main objective of this study is to develop and evaluate an online scale-selective nudging scheme for MPAS-A on an unstructured variable-resolution mesh. To make this clearer, we have revised the Introduction by adding a more explicit statement of the study goal around Lines 73–78.

Revised text: Building on the MPAS-A analysis nudging framework of Bullock et al., this study extends conventional full-field relaxation to scale-selective relaxation in MPAS-A version 8.2.2. To achieve this, we incorporate the diffusion-based filtering framework of Grooms et al. (2021), which enables efficient spatial filtering on unstructured meshes through repeated applications of the discrete Laplacian. Our goal is to constrain large-scale circulation toward the reference analysis while preserving

smaller-scale degrees of freedom in the refined region, and to evaluate whether the scale-selective nudging provides the intended scale-dependent nudging behaviour. This implementation provides a practical scale-selective nudging framework for variable resolution model, enabling parameterization sensitivity experiments under prescribed large-scale dynamical conditions while retaining the distinctive responses of different physics schemes that would otherwise be suppressed by full analysis nudging.

I think most of the content in section 2.2 is from Grooms et al. (2021), however, it is not well-organized. Please revise it carefully.

L105: I cannot find the Laplacian operator △ you have mentioned.

L126-127:" the diffusion operator is designed in the form of △"

I cannot find the diffusion operator you have mentioned.

Response: We thank the reviewer for pointing this out. We agree that the previous version of Sect. 2.2 did not present the method clearly enough and that the notation for the Laplacian and diffusion operators was confusing. We have therefore substantially revised Sect. 2.2, especially Lines 114–169, to reorganize the description of the diffusion-based filtering framework following Grooms et al. (2021) and to clarify how it is applied in MPAS-A.

In the revised manuscript, we now explicitly define the discrete Laplacian operator using the notation L. We also clarify that the diffusion-based filter is constructed through repeated applications of this discrete Laplacian operator, and we have rewritten the related explanation accordingly. In addition, the problems in Eqs. (5)–(8) have been corrected to make the mathematical formulation consistent with the revised notation.

L162: In the experiment design, in addition to the difference in convective parameterization, the PBL scheme and microphysics schemes are also different in

your experiment. Is there any reason not to use the same PBL and microphysics scheme? Can the author show some results to demonstrate that the convective parameterization plays a major role in contributing to the difference between the experiments?

Response: We thank the reviewer for this important comment. We agree that the original TK and GF experiments differ not only in the convective parameterization, but also in the associated PBL and microphysics schemes. In the original experiment design, we used these two configurations as internally consistent MPAS-A physics suites, rather than as single-parameter perturbation experiments.

We performed an additional 1-month sensitivity experiment in which only the convective parameterization was changed to GF, while the other physics options, including the PBL and microphysics schemes, were kept the same as in the TK suite. The daily precipitation frequency is now shown below. The green and red curves represent the original freerun_GF experiment and the new sensitivity experiment, respectively. The comparison shows that the precipitation-frequency distribution remains close to that of the original GF suite. This indicates that the GF convective parameterization plays a dominant role in shaping the precipitation-frequency characteristics discussed in this study.

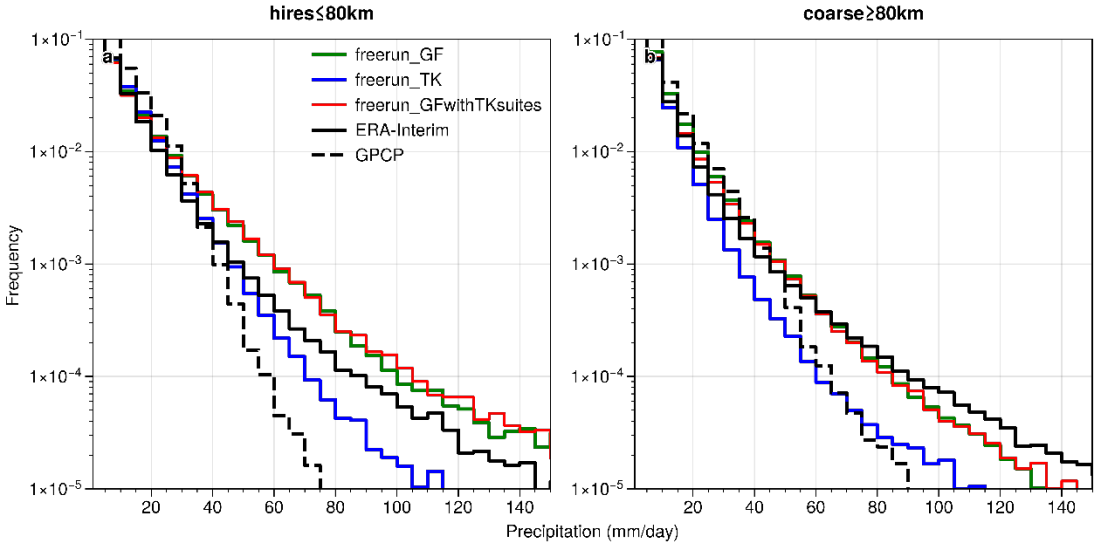

Figure 1. Daily precipitation frequency distributions comparing the original GF suite (green) and the GF-convection-only (red) sensitivity experiment.

L168: “altering the partitioning between parameterized and large-scale precipitation.” I’m curious about what the large-scale precipitation here refers to. The model is run with 92–25 km grid-spacing, so I assume all the convection is “parameterized”, is that correct?

Response: We thank the reviewer for this helpful comment. We agree that the original wording was not sufficiently precise. At the 92–25 km grid spacing used in this study, convection is not explicitly resolved, and deep convection is mainly represented by cumulus parameterization. Here, “large-scale precipitation” referred to the model diagnostic component produced by the grid-scale microphysics scheme, rather than explicitly resolved convective precipitation. To avoid confusion, we have revised the text around Lines 211–216. In the revised manuscript, we describe this more clearly as the partitioning between convective precipitation produced by the cumulus parameterization and grid-scale precipitation produced by microphysics. We also clarify that the scale-aware GF scheme can adjust the strength of convective parameterization with grid spacing, thereby modifying this partitioning.

In Table 2, the nstep values for gaussian1000km-uv_GF and gaussian1000km-uv_TK differ; is this the correct number?

Response: We thank the reviewer for carefully checking Table 2. We have re-examined the filter configuration and confirmed that the different nstep reported for gaussian1000km-uv_GF and gaussian1000km-uv_TK were a typographical error. Both experiments used the same Gaussian filter setting, and the correct nstep value is 20 for both cases. We have corrected Table 2 accordingly in the revised manuscript.

L225: I did not see any verification with IMERG product.

Response: In the original manuscript, IMERG was mentioned in the data description, but it was not used in the final verification after we revised the analysis. We have therefore revised the data description around Lines 281–287 and removed the statement related to IMERG. In the revised manuscript, the data section now only

introduces the observational products that are actually used in the verification.

L255: What is the observed standard deviation mean?

Response: We thank the reviewer for pointing out this unclear wording. Here, the "observed standard deviation" refers to the spatial standard deviation of the observational field over the evaluated region after applying the corresponding resolution mask. For each month, the model–observation RMSE is first calculated over the masked spatial domain and is then normalized by the spatial standard deviation of the observational field for the same month. We have revised the text at Line 314 to clarify this definition.

Revised text: The corresponding nRMSE fields are normalized by the observed spatial standard deviation of the observed field to provide a consistent metric across variables.

Results show that nudging with the analysis largely erases the scheme-dependent differences. I wonder if the nudging coefficient used here is appropriate or not? Have you tried different settings?

Response: We thank the reviewer for this important comment. The analysis nudging experiment is designed as a strong-constraint reference case, so the substantial reduction of scheme-dependent differences is expected and represents an approximate upper bound of the nudging effect. We tested different nudging coefficients for wind nudging. Compared with the coefficient of 3×10^{-4}, using 3×10^{-5}produced about 6 % differences in the spectral response: the wind field was overestimated at scales larger than the prescribed cutoff scale of 1000 km and underestimated at smaller scales. Increasing the coefficient to 3×10^{-3}caused unrealistically strong winds and model instability.

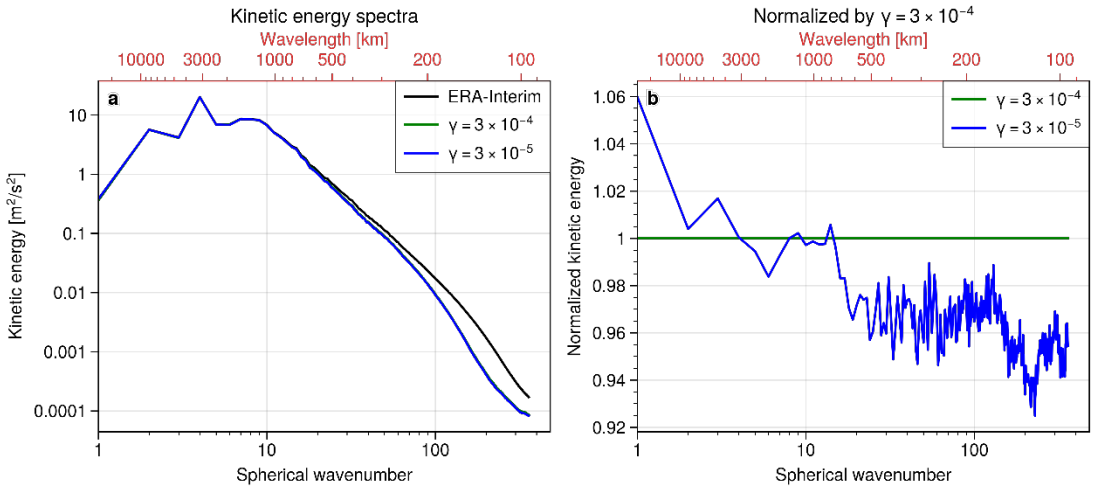

Figure 2. 500-hPa kinetic energy spectra. Left: spectra for the taper1000km wind-nudging experiments using different nudging coefficients. Right: normalized kinetic energy spectra, with each experiment normalized by the corresponding taper1000km case using a nudging coefficient of 3×10−4. This comparison illustrates how the nudging coefficient affects the scale-dependent kinetic energy response around the prescribed 1000 km cutoff scale.

L263-264: It is interesting that even with analysis nudging, the nRMSE in VORT500 remains large. Have you check the nRMSE in U500 and V500? Are they also not corrected by nudging?

Response: We thank the reviewer for this helpful question. We have additionally checked the nRMSE of U500 and V500. Unlike VORT500 in the following Figure 3, both wind components show robust improvement under analysis nudging, with nRMSE values generally between those at 200 and 850hPa. Therefore, the large VORT500 nRMSE does not indicate a failure of nudging to correct the 500hPa wind field. Instead, it suggests that VORT500 is more sensitive to internal mesoscale variability, especially over the refined region.

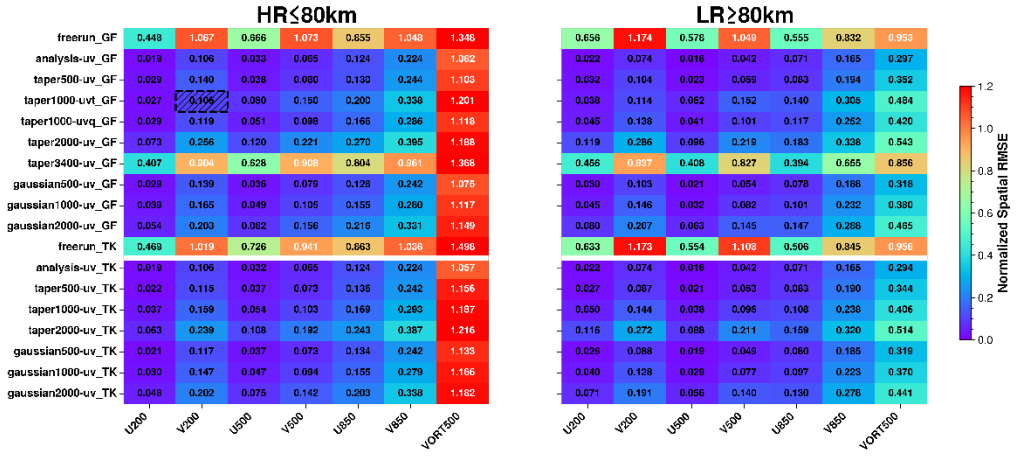

Figure 3. Normalized RMSE (nRMSE) over the refined region (≤80 km, left panels) and the coarse region (>80 km, right panels). The variables include 200 hPa 500hPa and 850 hPa wind speed (U200, V200, U500, V500, U850, V850), 500 hPa relative vorticity (VORT500).

L265-268: In GF, introducing potential temperature produces the worst precipitation nRMSE, while adding moisture nudging yields the best performance in precipitation. Is it the same in TK?

Response: We thank the reviewer for this helpful question. In this study, the additional uvt and uvq sensitivity experiments were mainly conducted for the GF suite, where precipitation frequency and thermodynamic–convective coupling show stronger sensitivity to the nudging configuration. For the TK suite, the available experiments suggest a relatively more robust precipitation response, and the improvement is mainly related to the wind nudging constraint. Accordingly, our discussion of the contrasting effects of temperature and moisture nudging is focused on GF. Further dedicated TK experiments would be useful to determine whether a similar response also occurs in TK.

L323: How to define “Nudging Efficiency”?

Response: We thank the reviewer for pointing out this unclear term. We agree that

nudging efficiency was not explicitly defined and could be misleading. In the revised manuscript, we have replaced this expression with scale-dependent nudging constraint, which more accurately describes the purpose of this section. We also added an introductory explanation at the beginning of Sect. 4.3, around Lines 390-393, to clarify that this analysis examines how strongly the nudging constrains different spatial scales and how much small-scale variability is retained under different filter settings.

Revised text: An efficient spectral nudging scheme should produce high model–reference coherence at wavelengths larger than the filter scale, while maintaining kinetic-energy and temporal variability close to the freerun at smaller scales. We evaluate this balance using kinetic-energy spectra, spectral coherence with ERA-Interim, and temporal amplitude spectra.

How do you compute kinetic energy spectra in this study? If you use 2D-FFT, do you need to interpolate simulation results to a uniform-distance grid? In such a case, what is your grid-spacing? Will it affect the result at a higher wave number?

Response: We thank the reviewer for this question. The kinetic-energy spectra are calculated from the global 500-hPa zonal and meridional wind fields using spherical harmonic decomposition. For each 6-hourly output, the spectra of the two wind components are combined to obtain the kinetic-energy spectrum as a function of spherical harmonic degree and then averaged over the full analysis period. We have added this clarification in Lines 395–399.

Revised text:

The kinetic-energy spectra are calculated from the global 500 hPa zonal and meridional wind fields using spherical harmonic decomposition. For each 6-hourly output time, the spectra of the two horizontal wind components were combined to obtain a kinetic-energy spectrum as a function of spherical harmonic degree. The resulting 6-hourly spectra were then averaged over the full analysis period for each experiment.

In Fig. 6, the taper1000-UV_GF and taper500-UV_GF have filter scales of 1000 and 500 km, respectively. But they separate at a wavelength of 400 km. What is the relationship between filter scale and wavelength defined here?

Response: We thank the reviewer for this question. The filter scale used in the experiment name refers to the nominal cutoff scale defined for the diffusion-based filter on the coarse-mesh background, rather than an exact spectral wavelength at which the kinetic-energy spectra must sharply separate. In practice, the damping response of the filter is gradual and can affect a range of smaller scales, especially after interaction with model dynamics and interpolation for spectral diagnostics. We have clarified this definition around Lines 253-258 and interpreted this around Lines 411-416.

Revised text:

This transition is better interpreted using the effective taper bands in the refined mesh. The relevant refined-region transition bands are 272–854 km for taper1000-uv and 136–427 km for taper500-uv. The taper1000-uv case begins to depart from the analysis-nudging response around 400 km, while the taper500-uv case shifts its main transition to a shorter wavelength, around 200 km; both wavelengths fall within their corresponding refined-region transition bands. These smaller transition wavelengths arise because the spectra combine contributions from both the coarse and refined parts of the variable resolution mesh, and the several-hundred-kilometer range is strongly influenced by the refined region response.

L345: Can you explain how the spectral coherence is computed?

Response: We thank the reviewer for this comment. We have added a description of the spectral coherence calculation around Lines 425-431. In brief, spectral coherence is computed from the cross-spectrum between the model and reference fields, normalized by the corresponding auto-spectra. It measures the scale-dependent phase and pattern consistency between the simulation and the reference analysis.

Revised text: Spectral coherence is used to measure the scale-dependent agreement

between the simulated and reference transient wind fields. We first remove the time mean of the 500 hPa horizontal winds and then compute the model self-spectrum, reference self-spectrum, and their cross-spectrum at each spherical harmonic degree l. The coherence is calculated as:

$$\rho_l = \frac{\sigma_{xy}(l)}{\sqrt{\sigma_{xx}(l)\sigma_{yy}(l)}}, \quad (9)$$

where $\sigma_{xx}(l)$ and $\sigma_{yy}(l)$ are the time-mean spectral variances of the simulated and reference transient winds at spherical harmonic degree l, respectively, and $\sigma_{xy}(l)$ is their time-mean spectral covariance.

In Fig. 7c and 7d, why do all experiments have a minimum at a wavelength of around 120 km?

Response: We thank the reviewer for this question. The common minimum around 120 km is mainly related to the effective resolution of the reference data. In this study, ERA-Interim is used as the reference for the coherence calculation, and its native horizontal resolution is about 0.75°, so scales below approximately 120 km are largely affected by interpolation and should not be over-interpreted. The comparison is therefore more meaningful at wavelengths larger than about 120 km. We have clarified this point in the revised manuscript around Lines 443-445. We still retain the smaller-scale part in the figure to show the relative behavior among freerun, analysis nudging, and different scale-selective nudging cases.

Revised text: The common minimum around 150 km is mainly related to the limited effective resolution of ERA-Interim (0.75°); below this scale, the spectra are strongly affected by interpolation, while the enhanced amplitudes in all simulations more likely reflect internally generated perturbations over the refined 25km mesh.