Reviewer #1

1. Detailed comparisons of the simulations against radar observations have been conducted for hail presence within storms using both the original and modified schemes. However, it is also crucial to examine the corresponding surface hail distribution for both rain and hail events, as accurate surface hail prediction, including hail size and amounts are vital for operational weather forecasting.

Reply: Thanks for your advice. We confirm that no hailstones were observed in the three heavy rainfall cases of this study. For the three hailstorm cases, the maximum hailstone diameters observed on the surface were 3–5 cm (Hailstorm 1), 5–8 cm (Hailstorm 2), and 5–6 cm (Hailstorm 3), respectively. Since our ground hail observations were not obtained from meteorological Bureau but rather from short video platforms like Douyin (TikTok), we only have information on hailstone diameters but lack data on hail amounts. Hail size information is shown in Section 2.3 in the original manuscript.

2. According to dual-polarization radar observations using the HCA, no hail was identified throughout the entire lifespan of the three heavy rainstorms. These findings contradict the simulated rainstorms, which showed a large amount of hail particles even with the modified Milbrandt scheme. Please provide further details on this inconsistency.

Reply: In the model, we defined three criteria for hail identification to ensure that the detected hail signals genuinely represent hail rather than spurious artifacts caused by size sorting: (1) the hail mass mixing ratio must exceed 0.1 g kg^{-1}, (2) its mass-weighted mean diameter must be greater than 2 mm and larger than that of graupel at the same model grid point, and (3) its equivalent reflectivity factor (Z_e) must exceed the combined Z_e of all other hydrometeor species (including cloud, rain, ice, snow, and graupel). The absence of detectable hail when applying these criteria with the modified MY2 scheme does not imply a near-zero hail mixing ratio (q_h), as the model still contains regions with hail content below 0.1 g kg^{-1}, mass-weighted diameters smaller than 2 mm or reflectivity values lower than those contributed by other hydrometeors. However, these extremely weak hail signals can be considered insignificant for practical purposes. The following is added in the revised manuscript: "*The simulations using the modified MY2 scheme do contain regions with positive hail mass mixing ratios (>0 g kg^{-1}) that fail to meet the three specified hail identification criteria. However, these regions exhibit hail signals below the detection thresholds and are therefore negligible*".

3. The graupel-to-hail conversion in the revised scheme requires a substantial amount of supercooled water collected by graupel at one model time step, specifically at least 1.25 times the mass of graupel particles. This criterion may be too stringent, which explains why it mitigates hail overprediction within rainstorms. However, while hail can be produced within the hailstorms, the simulated radar reflectivity and hail distributions are underestimated compared to radar observations. Additionally, the integration time step also influences the graupel-to-hail conversion, which is also not physically reasonable.

Reply: You are right. Due to the fixed graupel and hail density settings in the MY2 scheme, graupel particles must accrete a substantial amount of supercooled water to fill their internal voids—a quantity exceeding the graupel's own mass. This physical constraint inherently inhibits hail formation and likely explains why the maximum radar reflectivity values in hail cases simulated by the modified MY2 scheme are lower than observed values. However, this discrepancy stems fundamentally from the MY2 scheme's predefined density settings rather than inaccuracies in our implemented spongy wet growth formulation. Were the spongy wet growth process incorporated into a prognostic density scheme like the NSSL scheme (which predicts graupel and hail densities), graupel particles would require significantly less supercooled water accretion for hail conversion. Consequently, the simulated reflectivity patterns would better match the observation.

I concur with your suggestion regarding the influence of the time step (*dt*) on hail conversion. Excessively large dt values may lead to $(Q_{rg}+Q_{cg})*dt \geq Q_{gwet}*dt + Q_{fill}$, resulting in artificially enhanced hail formation. Therefore, implementing an upper limit for *dt* through comprehensive sensitivity experiments is needed in future studies. However, given our current configuration with 1-km horizontal resolution and 2-s time step, maybe the *dt* is sufficiently small to prevent this overestimation of hail production. The following is added in Discussion: "*Our proposed method of introducing spongy wet growth to improve the graupel-to-hail conversion process still requires further investigation in several aspects. For instance, the model integration timestep may exert an influence on the conversion occurrence. Because larger dt values increase the likelihood that the supercooled water content accreted by graupel particles exceeds the amount required to fill their internal voids, thereby artificially enhancing hail formation. Although our current experiments employ a sufficiently small dt (2 s), comprehensive sensitivity tests are needed in the future to provide optimal dt thresholds. Furthermore, the predefined graupel and hail density in the MY2 scheme physically constrain hail conversion by requiring graupel particles to accrete sufficient supercooled water within a single timestep. This inherent configuration may systematically suppress hail production. Future implementation of this approach within microphysical schemes which could predict the density of graupel and hail particles, such as the prognostic-graupel-density MY2 (Milbrandt and Morrison, 2013) and the NSSL scheme, could potentially mitigate this limitation. Additionally, the absence of high-density ice particles above the freezing level, as shown in the P3 scheme (Morrison and Milbrandt, 2015; Milbrandt and Morrison, 2016; Milbrandt et al., 2021) by Labriola et al. (2019b, Fig. 14a in their study) and the Thompson_Hail scheme (Jensen et al., 2023) by Li et al. (2024, Fig. 8f in their study), could potentially be improved by introducing the spongy wet growth process to increase the density of riming particles*".

4. The discussions in the last two paragraphs of Section 3 lack clarity. For instance, while both thermal buoyancy and dynamic vertical pressure gradient force are important for the vertical motion, only the buoyancy term is investigated.

Reply: Thanks for your comments. We have calculated the dynamic vertical pressure gradient $-\frac{1}{\bar{\rho}}\frac{\partial p'_{\mathrm{d}}}{\partial z}$, where p'_{d} is dynamic pressure and is calculated as:

$$\nabla^2 p'_{\mathrm{d}} = \bar{\rho}(|\,\boldsymbol{\omega}\,|^2 - |\,D\,|^2)$$

$$|\,\boldsymbol{\omega}\,|^2 = (w_y - v_z)^2 + (u_z - w_x)^2 + (v_x - u_y)^2$$

$$|\,D\,|^2 = (u_x^2 + u_y^2 + u_z^2) + (v_x^2 + v_y^2 + v_z^2) + (w_x^2 + w_y^2 + w_z^2) - (\mathrm{d}\ln\bar{\rho}/\mathrm{d}z)\boldsymbol{v}\cdot\nabla w - (\mathrm{d}^2\ln\bar{\rho}/\mathrm{d}z^2)w^2$$

where $\boldsymbol{\omega}$ is the three-dimensional vorticity vector, $|D|$ is the magnitude of the total deformation, $\bar{\rho}$ is the basic state density. Its magnitude ranges between approximately -0.03 to 0.03 m s^{-2}, which is an order of magnitude smaller than the buoyancy force (-0.2 to 0.2 m s^{-2}). This indicates that the dynamic pressure has a negligible contribution to vertical acceleration compared to buoyancy. Furthermore, the strong spatial correlation between updrafts and positive buoyancy regions (Figure 12) further confirms that vertical motion is primarily driven by buoyancy. The following is added in the revised manuscript: "*Given that the dynamic vertical pressure gradient force (about 0.03 m s^{-2}) is about an order of magnitude smaller than the buoyancy force (about 0.2 m s^{-2}) in these six cases (not shown), our study primarily focuses on the dynamical effects of buoyancy*".

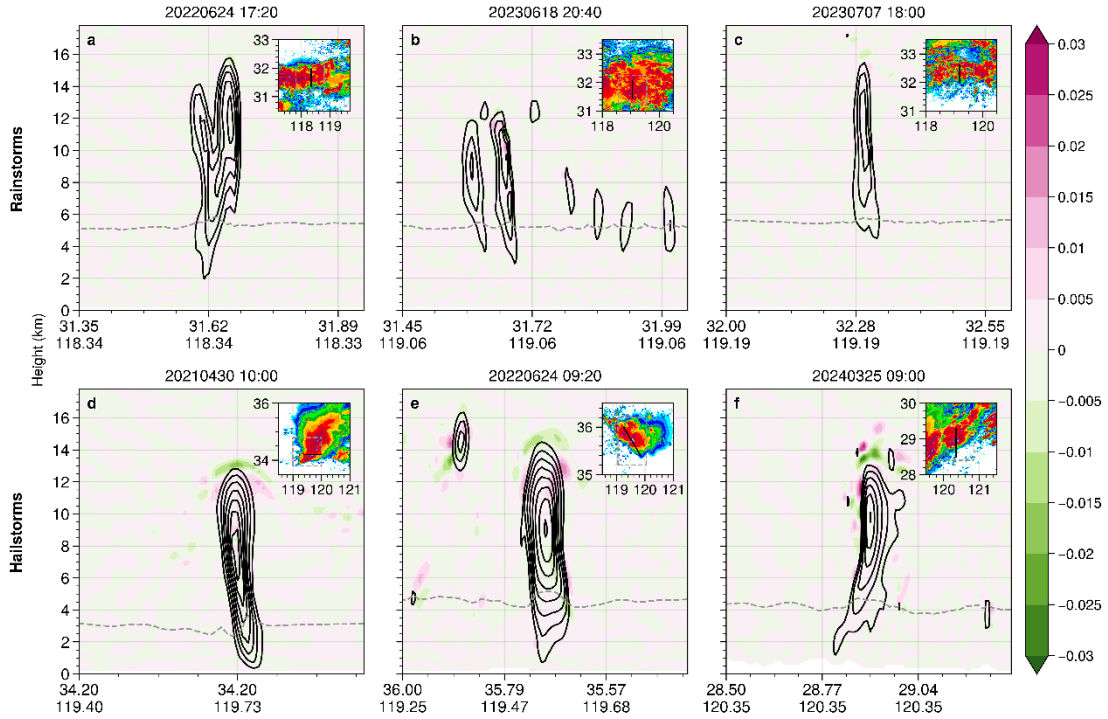

Figure R1. Similar to Figure 12 but for dynamic vertical pressure gradient force (color shading, unit: m s^{-2}). Vertical velocity (black lines) contours range from 5 to 40 m s^{-1} at intervals of 5 m s^{-1}.