Reviewer #1: This is a detailed study using WRF and 4 microphysics schemes for 8 snow events during ICE-POP for the 2018 Winter Olympics, focusing on one of each of 3 types of events, cold-low, warm-low and air-sea interaction. The inner domain is at a relatively high resolution of 1 km. Observations used in addition to the surface AWS stations include disdrometers and radar from which particle types were derived. The paper distinguishes some large differences in particle types between the schemes and verifies them against observations. Some additional understanding is gained by evaluating the importance of every process as a source and sink for each particle type in each case. This is a large amount of data that is presented, and a good attempt is made to derive the most important points and distinctions between the schemes from it. I think the paper is acceptable after minor revisions. The level of detail may appeal mostly to microphysics parameterization developers, and is probably more than most would read through, but the conclusions are of more general interest. I have itemized my Minor Points below, the response to some of which may help to improve the paper.
: We appreciate the valuable reviews. The manuscript has been revised in accordance with reviewer's comments and suggestions. Please find the answer for each comment below.

Minor points
1. line 33. What is meant by "inefficient melting"? Less efficient?
: The corresponding phrase has been revised as "by reducing the melting efficiency in all schemes".

2. L52. convections -> convection here and several places. Common English error.
: Revised accordingly throughout the manuscript.

3. L76 Thompson.
: Revised accordingly.

4. L76 "snow efficiently affects precipitation efficiency for" rephrase to not include efficient twice.
: In response to reviewer's comment, the word, efficiently" has been deleted.

5. L114 "of precipitationi systems" typo.
: Revised accordingly.

6. Table 2. Refers to Morcrette. I am sure this is not the correct reference.
: In response to reviewer's comment, we have deleted that reference in Table 2.

7. Table 4. WDN typo.
: Revised accordingly.

8. L213 and Figure 5. Case color code should be mentioned in the text too.
: In response to reviewer's suggestion, we have added the following sentences in the revised manuscript.
"White, black, yellow, and blue-colored bars represent the results for the simulations with the WDM6, Thompson, and Morrison schemes. The cold-low, warm-low, and air-sea interaction cases are shaded in blue, red, and green color."

9. L216. How is a rate used for an accumulated amount in the whole period? It says mm h^{-1}.
: For the total cumulative precipitation [mm], any threshold value for the rain rate is not adopted.

10. L218 and Figure 5. Hard to interpret biases without absolute totals which vary from 6 mm in Case 3 to 57 mm in Case 4. Perhaps put totals from Table 1 on Figure 5.
: In response to reviewer's suggestion, we have added the total precipitation amount in Figure 5a and modified the caption.

"Figure 5. ... The total cumulative precipitation [mm] for each case, obtained from the AWS (Table 1), is also noted in Figure 5(a) using red dots together with the scale in the right y-axis."

11. L222. Hard to tell from Figure 6 that these schemes have more liquid rain. I would suggest finding a different way to show precip type. Either a separate plot of type, or shading by type and contouring amount.
: We have added the contour lines presenting the rain-type precipitation in the Figure 6. Accordingly, the manuscript and figure caption have been modified as below.
The following sentences, "All schemes simulate the precipitation as a type of snow <u>and rain</u> over the northeastern part of the domain. WDM6 and WDM7 simulate more liquid rain at the surface precipitation than Morrison and Thompson.", are modified to ""All schemes simulate the precipitation as a type of snow over the northeastern part of the domain."
"Figure 6... <u>Black,</u> red, blue, and <u>purple</u> contours represent the <u>rain,</u> snow, graupel, and hail-type precipitation at the surface. <u>The contour intervals for CASE 3, CASE 6, and CASE 7 are 3, 5, and 10 mm.</u>"

12. Figure 8. It was hard to find qc because the dash length does not match the key. Make the key pattern exactly match the plot. Also hard to see that qs is a dot pattern in the key.
: In response to reviewer's comment, we have modified the figure to make the key pattern match with the plot.

13. L262. Important to note that schemes with QCGEN have condensation mostly there while those without combine condensation and evaporation in QCCON. Is there much separate condensation in QCCON in the QCGEN schemes or is this all just evaporation?
: To deliver the results clearly, we have added the following sentence in the revised manuscript.
"QCGEN includes only the condensation, but QCCON includes both condensation and evaporation. The negative sign of QCCON means that the magnitude of evaporation is greater than that of condensation."

14. Figure 9a-d. Maybe QRWET should be QCWET in labels. Check all these against Table 4 names.
: In response to reviewer's comment, the notation of QCWET has been changed to QRWET in Table 4. In addition, other notations have been checked again.

15. Table 4. QRAUT in cloud section could be QCAUT? I am not sure about the rules for naming when the same processes may have different names. QCACR for example has the same name.
: The identical microphysical processes have the same notation. QRAUT is the source of rain, but the sink of cloud water. This is same for QCACR.

16. Figure 9, etc. Can a scaling number be put on these plots to show relative size? L270 points out an important scale difference that would not have been seen in the Figure. For example, add what 100 equals in absolute terms.
: In response to reviewer's comment, we have added the scaling number in Figures 9, 11, and 13 in the upper left corner. In addition, the figure caption has been modified as below.
"Figure 9. Relative contribution of time-domain averaged production <u>tendency</u> term during the analysis period. From the left column, figures indicate the simulation results with the WDM6, WDM7, Thompson, and Morrison schemes. (a)–(d) are the terms for cloud water, (e)–(h) for rain, (i)–(l) for cloud ice, (m)–(p) for snow, and (q)–(t) for graupel, and (u) for hail. The hail is only predicted in WDM7. <u>The scaling number, sum of the absolute value of each production tendency, which corresponds to 100%, are noted in the upper left corner of each figure.</u>"

16. L301. As in note 11 above, this is hard to see.
: As noted in our response for the comment #11, we have added the contour lines presenting the rain-type precipitation in the corresponding figures.

17. L372. Should be Fig. 7l.

: Revised accordingly.

18. L373. Westerly wind is weak. The model clearly has an onshore wind that must be mainly northerly. This component should be mentioned.

: Thank you for the comment. When we investigated the u and v wind components over the cross-sectional area in all simulations, it was confirmed that the v component is smaller than the u component. As we noted in the original manuscript, the wind is mostly westly.

Reviewer #2: This study evaluates the performance of four bulk microphysics schemes for the simulation of snowstorms during the Pyeongchang 2018 Olympics and Paralympic Games. The analysis compares the amount of ice, snow, precipitation and cloud water predicted by the various schemes, and attributes the processes giving the production of the relevant hydrometeor categories. They conclude that melting is key for generating rain, and the bias in precipitation for war-low and cold-low biases can be mitigated through the use of inefficient melting in all schemes. Although I think the paper makes a contribution by testing these microphysical schemes in new meteorological situations, there are aspects of the presentation that should be improved before the paper is accepted for publication.

: We appreciate the valuable reviews. The manuscript has been revised in accordance with reviewer's comments and suggestions. Please find the answer for each comment below.

MAJOR COMMENTS

1. I found that the introduction was not overly focused. Although the stated goal of this study is to evaluate the performance of the microphysical schemes in the simulation of wintertime precipitation, much of the introduction compared how these schemes have previously been used to simulate convection. There should have been more focus on how the use of these schemes has been evaluated in simulations of wintertime storms, and also the understanding that has been gained from past observational and modeling studies of winter storms should be highlighted. There were many past studies of winter storms that were not referenced.

: In response to reviewer's comments, we have modified the introduction to address additional evaluation results of microphysics schemes in simulations of wintertime storms. The following references, Solomon et al.(2009), Molthan and Colle (2012), Conrick et al. (2019), Ma et al. (2021), have been newly cited, and the understanding from these studies has been noted as below.

"Through the modeling and observational studies of winter storms, the major microphysics processes affecting the characteristics of winter storms have been figured out (McMillen and Steenburgh 2015; Lim et al., 2020; Ma et al. 2021) and the cloud microphysics parameterizations have been evaluated by utilizing the measurements from extensive observation campaigns (Solomon et al. 2009; Molthan and Colle, 2012; Conrick et al. 2019)." "....Ma et al. (2021) emphasized that the cloud ice deposition/sublimation parameterization greatly affects to the snowfall amount. By altering this parameterization in WSM6 scheme, the overestimation of the snowfall amount was notably reduced in WRF simulations."

"Solomon et al. (2009) verified the microphysical characteristics for the simulated mixed-phase clouds by utilizing the intensive measurements taken during the Mixed-Phase Arctic Cloud Experiment (M-PACE). They showed that the double-moment microphysics scheme simulates realistic liquid water paths, compared to the single-moment scheme. Through the comparison between the observation data during The Canadian CloudSat/Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) Validation Project (C3VP) and assumptions used in microphysics schemes, Molthan and Colle (2012) concluded that single-moment schemes having a flexibility in

size distribution parameters as functions of temperature can represent the vertical variability as observed ones from aircraft data. Conrick and Mass (2019) evaluated Thompson microphysics scheme in the WRF model using observations collected during the Olympic Mountains Experiment (OLYMPEX) field campaign by the Global Precipitation Measurement (GPM) satellite and showed that Thompson scheme underpredicts radar reflectivity below 2 km and overpredicts one above 2 km, consistent with the vertical mixing ratio profiles from GPM Microwave Imager."

2. I am concerned about the resolution of the model that is being used as the highest horizontal grid spacing is 1 km. A lot of the convection and generating cells that commonly occur in winter time storms are on scales of much less than 1 km. Thus, the model is not able to represent well the spatial scales on which the evolution of these storms is occurring. This limitation should be clearly acknowledged (or run at finer resolution) and explain the caveats with the interpretation of this study because of this resolution difficulty.

   : We appreciate the reviewer's comment. We have added the following sentences in the "Summary" section of the revised manuscript.

   "Even though several studies simulated snow storm cases under the horizontal resolution of 1-km or 1.33 km (Alcott and Steenburgh 2013; Molthan et al. 2016; Vigonon et al. 2019; Veals et al. 2020), the 1-km horizontal resolution, used in our study, could be coarse for some generating cells during winter season. However, these small-scale cells cannot alter the major findings of our study."

3. The authors use four different microphysical schemes in their investigations, but do not use some of the most state-of-art microphysics schemes in their simulations. Why is this? For example, the P3 schemes (Morrison and Milbrandt 2015), Predicted Particle Properties scheme is the next generation parameterization scheme that uses a very different approach for representing ice, and would offer an interesting complement to the schemes that are presented here. It predicts bulk properties rather than predicting separate species, which eliminates unphysical conversion processes between traditional ice categories, and hence can be used for giving a better comparison against observations.

   : We agree that the P3 scheme can offer an interesting complement to other microphysics schemes utilized in our study. However, the criteria for dividing ice particles into snow and graupel are quite arbitrary in P3. Therefore, the comparison of each solid-phase hydrometeor between P3 and other microphysical schemes is not straightforward. This is because P3 predicts the total solid-phase (cloud ice + snow+graupel) mixing ratio and rimied-particle mixing ratio, differently from the conventional microphysics schemes such as Thompson, Morrison, and WDM6(7). We have tried to evaluate the simulated mixing ratio using P3 scheme (Fig. R1). The simulated amount of graupel and snow mixing ratio can differ depending on the threshold value of rimed fraction, which is calculated using QIR/QICE (rimied-particle mixing ratio/total solid-phase mixing ratio) (Fig. R1).

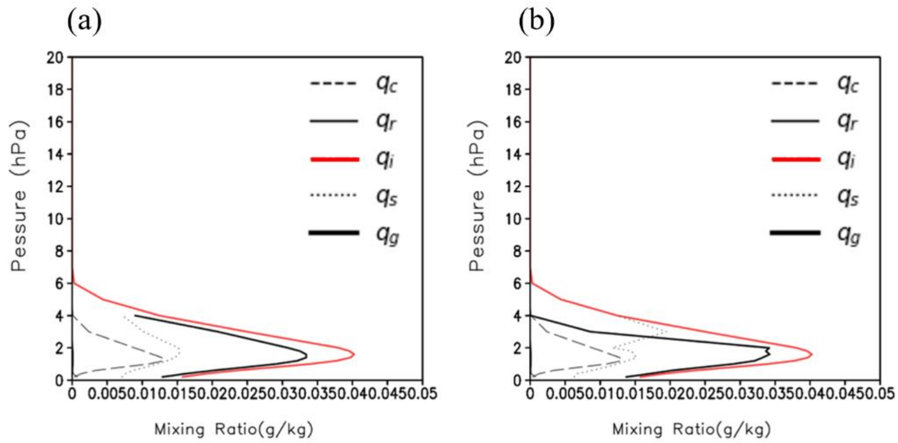

Figure R1. Time-domain averaged vertical hydrometeor mixing ratio profiles from the P3 simulations when the ratio of QIR/QICE is 0.3 (left) and 0.8 (right). The averaged time period and domain are the same as CASE 3 in Figures 7.

4. For the most part, the writing in the manuscript where the results of the different simulations and comparisons against observations is performed is overly qualitative. Terms such as "overestimate", "matches well", "substantial amount", "abundant", "insignificant", "similar" and others are used excessively, with little information on what this actually means. The paper would read more clearly if these descriptions were made more quantitative with appropriate reference to the figures. There are a few places where this is done (e.g., approximately 10 times larger, magnitude is 5.5 g/kg), but this could be done much more effectively.

: In response to the reviewer's comment, the qualitative expressions in the result section of the original manuscript have been revised. The sentences, shown below, are the part of examples.

L 243: "All microphysics parameterizations overestimate the surface precipitation amount." → "All microphysics parameterizations present a positive bias for the surface precipitation."

L 258: "Furthermore, cloud ice is predicted, even near the mountain top, with a substantial snow amount." → "Furthermore, cloud ice is predicted, even near the mountain top, with a snow amount greater than 0.38 g kg^{-1} at around 1.5-km level.

L 266: "As shown in Figures 8a and b, the vertical distributions of hydrometeors from WDM6 and WDM7 are similar, except hail." → "As shown in Figures 8a and b, the vertical distributions of hydrometeors from WDM6 and WDM7 are comparable in terms of the vertical extent and the maximum level of hydrometeors, except for hail."

L 328: "However, WDM7 simulates abundant hail-type precipitation over the southeastern part of the analysis domain." → "However, WDM7 simulates hail-type precipitation amount more than 10 mm over the southeastern part of the analysis domain."

L 355: "The major sinks and sources of the liquid hydrometeors are similar." → "The major sinks and sources of the liquid hydrometeors are identical.

L 359: "…, leading to an abundant production of cloud ice." → "…, leading to an abundant production of cloud ice greater than 0.06 gkg^{-1} (Fig. 10ab)."

L 383: "WDM6 and WDM7 simulate abundant solid-phase precipitation compared with the simulations with Thompson and Morrison." → WDM6 and WDM7 simulate solid-phase precipitation amount more than 14 mm."

L 409: "Both schemes simulate abundant cloud ice compared to Thompson and Morrison in CASE 7 as well. " → "In WDM6 and WDM7 schemes, the magnitude of QIDEP is 0.27 g kg^{-1}, which is about 10 times larger than that in Thompson and Morrison."

L 415: "…, but the formed graupel amount is insignificant." → "…, but the formed graupel amount is not identified in the surface precipitation."

L 433: "The retrievals from the radar also show abundant cloud ice and snow over the downslope region of the mountain top." → "The retrievals from the radar also classify cloud ice and snow as primary hydrometeor types over the downslope region of the mountain top."

L 443: "WDM7 predicts abundant hail-type precipitation." → "WDM7 predicts hail-type precipitation amount more than 10 mm, which is not observed."

L444: "WDM6 and WDM7 simulate the abundant cloud ice.." → "WDM6 and WDM7 simulate cloud ice amount between 0.01 and 0.1 g kg^{-1}"

L 447: "Meanwhile, Morrison and Thompson simulate abundant snow over the corresponding region." → "Meanwhile, Morrison and Thompson simulate more snow over the corresponding region, compared to WDM6 and WDM7."

L 450: "Through the microphysics budget analysis, it is found that the major sources and sinks of hydrometeors are similar between the cold-low and warm-low cases." → "Through the microphysics budget analysis, it is found that the major sources and sinks of hydrometeors are identical between the cold-low and warm-low cases."

L 459: "…, simulations for the air-sea interaction case produce abundant cloud water." "…, simulations for the air-sea interaction case produce abundant cloud water amount greater than 0.2 g kg^{-1}."