Review of "The impacts of secondary ice production on microphysics and dynamics in tropical convection", by Qu et al.

This paper investigates the importance of two SIP processes on tropical convection using high-resolution simulations. They also make use of observations from the HAI-HIWC program to provide some evaluation of the performance of the simulations.
I really enjoyed reading this paper, which is very well written and organized. I have a few comments, mostly regarding the comparisons with the observations, which would need to be addressed before the paper is ready for publication. These comments are listed below.

Main comments

- In the paragraph starting line 299, the authors comment that the differences in IWC statistics at 11-12 km could be due to differences in the sampling of data, with the HAIC-HIWC program targeting conditions with high IWC. However, the same comment applies to all comparisons in this paper, including those that seem to work like IWC at 6-7 km, and number concentrations. I think it is the main weakness of the paper not to have attempted to minimize these differences in sampling. There are different ways to address this problem. What I suggest would make more sense is to compare relationships between parameters. For instance, Ni versus reflectivity, IWC versus reflectivity, at different heights (6-7 and 11-12km like you've done is good) so that you can compare in different ranges of reflectivity how the Ni and IWC distributions compare with the observed ones. Or use vertical velocity like you've done later in the paper when looking at SIP rates. I am very confident that with a little more care you will be able to say more about how well the simulation captures what the observations say.
- Results from Fig. 6, especially IWC and Z: This is a good illustration of my previous point, Differences in IWC and Z should result in very different IWC-Z relationships from the two simulations. You could compare simulations in IWC – Z space and compare with observations of IWC and Z (simulated from PSDs or measured by NAX close to the aircraft) from HAIC-HIWC too. Regions of simulated storms that were not measured by the aircraft will not be covered by observations, but that's OK, just compare where you can in the range of possible IWC and Z values.
- Fig. 12: The Convair flew slightly above or slightly below the 0degC isotherm. If not corrected for attenuation, your stats from the Convair cannot be used as a reference. Has attenuation been corrected? I suspect not. A maximum of radar Z in liquid of ~25 dBZ with a maximum at ~ 40 dBZ at the right side of the CFAD seems too low for tropical convection at X-band.
- Comparisons from Fig. 12: I think there is more to say about this figure. I also think the colour scale of your reflectivity CFADs is not well chosen it does not highlight enough the maxima that are of most interest:
  - just below the melting layer there is a local maximum of rainfall in the SIP simulation at ~ 25 dBZ, where the observations show a maximum. It corresponds to a blob of higher frequencies ion the CFAD produced by SIP around +15 dBZ above the melting layer. It is very impressive! Probably the most important result I see in those comparisons.
  - The SIP simulation still produces very large ice particles (30 dBZ+) in the HM region that are completely absent from the observations. That is a long-standing problem in tropical convection simulations, it needs to be mentioned and discussed. Used to be associated with way too much graupel produced by the models. But in your case you don't use the same type of parameterizations, so what causes this population of ice particles to be there ? Calls from a deeper analysis of your four ice categories ? Is there one in particular that

occupies this space on the reflectivity CFAD ? How are the four ice categories spread out on the CFAD ? That would be ar eally interesting plot to make and analyze, maybe.

- Following up from the previous comment, nothing is done to show how the four ice categories end up looking like and how they contribute to the total Ni, IWC, etc ...

Other comments:
Line 168: " There have BEEN" ("been" is missing I think)

Line 177: Mossop (not Mossp)

Section 4, lines 239 - 240: it would be interesting to document somehow what difference this choice makes to the morphology of the storm for SIP simulations. Would it be possible to compare cross sections (horizontal or vertical) as well ? Or comment on how much that choice changes the morphology of the storm without showing figures? Seems like a good opportunity.

Figure 7: The only parameter that shows much less difference between simulations in the ensemble exercise is vertical velocity. Any idea why ? You should offer more explanations about that, and acknowledge that as well.

Lines 285-288: Here you need to talk about the width of the distribution. Although the max in obs is incidentally close to the CTR simulation, the width of the distribution is much better captured by the SIP simulation. Then you need to investigate why there is this high population of 10^6 Ni in the simulation, maybe with some sensitivity tests ? If it's good at 6-7 km, but not at 11-12 (higher concentrations), it seems unrelated to how well you simulate SIP maybe?

Section 5.4 : differences in LW radiation are really interesting and show substantial differences between non-SIP and SIP comparisons. Why didn't you attempt to compare with satellite products?

Lines 367-368: If I understand correctly panel a is not normalized by the actual areas of the three updraft categories. Maybe the lower values for downdrafts is just because there is much less area actually occupied by downdrafts? I wonder if it would be complementary to look at an activation fractional area (% of downdraft area where SIP is doing something)?

Lines 392-395 and line 427: This reads like you conclude it's like that in real life tropical convection. Most of this is driven by the choices you made for the parameterization, so maybe this comment should be restricted to "in these simulations ...". The same caution should be exercised in the general conclusions, I think.

Line 451: We sort of discover here that your objective is to test whether a higher collection efficiency has large influence on the result. I think you should state clearly in lines 444-448 that you are taking an extreme case of higher efficiency to explore. This also calls for another question: why did you chose the default one you used? Is it known to be more physically realistic?

Fig. 20: it is striking that the vertical velocities are a lot lower at 1km compared to 250m. This should be mentioned I guess? Also I think it would be interesting to add the profiles of Fig. 6 (as dashed line) on Fig. 20 to support this discussion.

Good luck with the Review
Alain Protat
Melbourne, 20/06/2022