Review of "The Chance of Freezing – Parameterizing temperature dependent freezing including randomness of INP concentrations" by Frostenberg et al.

General comment:

This study presents a new parameterization for heterogeneous cloud ice nucleation, producing INP concentrations as a function of temperature only. Previous parameterizations of INPs as a sole function of temperature (such as the one of Fletcher, 1962) produce a single value of INP concentration for a given temperature T, in order to reproduce the median INP=f(T) relationship observed globally and the overall temperature dependence of cloud ice nucleation. However, the approach of Fletcher and similar models fail to reproduce the very wide range of INP observed globally at a given temperature. The parameterization proposed in this study uses a stochastic approach, by including a range of INPs that can be sampled randomly from a frequency distribution around the central INP=f(T) value, so that a sufficiently large sample from the parameterization reproduce the observed range.

I found this idea interesting and worthwhile to explore. The authors have presented their results clearly and with a lot of care to explore sensitivities and impacts. Unfortunately, unless I am very mistaken, the results presented in the manuscript clearly show that this approach does not work when coupled to a cloud model and that it should not be used in such models, so I cannot recommend it for publication in ACP. Details are given below under "Major comments".

I realize that a large amount of work has been done for this study, and I actually found that understanding why this approach fails was very illuminating for understanding how INP parameterization works and their effects on clouds (see for example the comments on lines 231 and 260, and Major comment 1). I also believe such negative results are very useful for future research and should not be harder to publish than positive results (see for ex. Matosin et al., 2014). For this reason, I think that this study still be valuable to publish with a major overhaul of the manuscript and a change of title to clearly present these teachings of this experiment.

Major comments:

To clarify, I think that there are 4 major issues with the approach.

1) There is I think a fatal flaw in the stochastic approach of the parameterization, which is apparent in the high sensitivity of your results to the drawing frequency, and in the divergence of IWP at high drawing frequencies (Section 3.1.4 and Figure 9). If the drawing frequency is increased, the INP concentration should be more representative of the observed INP variability, which should lead to more accurate results if the parameterization works as intended.
   On the opposite, results in Figure 9 show that increasing the frequency leads to arbitrarily high IWP, which does not even converge toward a given value at higher frequencies. On the opposite, the IWP actually converges for low drawing frequencies, when drawn INPs are less representative of the underlying distribution. Figure 9 only shows frequencies down to 0.2 Hz. However, with the study's approach, I am convinced that increasing the frequency further would produce arbitrarily high ice formation, and a high enough frequency would cause 100% freezing of all supercooled cloud droplets.
   The reasons for this behavior are partly discussed in the text, and to clarify I will present my understanding here. Large INP concentrations in the atmosphere lead to cloud ice nucleation if the thermodynamic conditions for ice formation are present: for deposition freezing, if the air is

supersaturated with respects to ice; for immersion and contact freezing, if there are supercooled liquid droplets present. Once formed, ice crystals are thermodynamically stable (as opposed to supersaturated air or supercooled liquid) and they do not disappear if INP concentrations decrease. As briefly mentioned in the manuscript, his “INP memory” of ice crystals explains the sensitivity to both the standard deviation of the frequency distribution and the drawing frequency. Since the lognormal distribution is unbounded, any sufficiently large sample from the distribution will contain any arbitrarily large value of INP concentration INPC>Nc that can cause total supercooled droplet freezing. High drawing frequencies correspond to a larger number of samples that are more likely to contain high values, and large standard deviation means that drawing a large INP value is more likely for a given sample size.

Unfortunately, this is not only due to the unbounded lognormal distribution. Using a bounded frequency distribution would solve this issue either, because then a high enough drawing frequency would produce INPC values from the upper bound of the distribution (instead of indefinitely large values for the lognormal case). This would be equivalent to a Fletcher-like parameterization, but using this arbitrary new upper bound of the distribution instead of the median, with, in my opinion, no added value from the stochastic approach, but an additional complexity, increased computational cost, and lower reproducibility.

Conversely, do not I think using a very low drawing frequency makes sense. As mentioned above, it would mean that less representative INP concentrations are required for the parameterization to work, which defeats its whole purpose of better representing the observed INP range. In addition, using a very low frequency means that in a given grid cell the INP concentration is random (possibly very high or very low) but does not change at all over the drawing period, which is also very unrealistic. This would mean that a single grid point in the model domain will for example always freeze all droplets (or freeze none) during the whole frequency time, which is much larger than the advection time scale, during which different air masses or clouds can pass through this permanently high-INP (or low-INP) grid cell.

2) The parameterization represents the global and annual INP variability for a given T, modeled as a random process, and reproduces this variability at all times and locations. However, I think that this observed variability arises largely from non-random local (temporal and spatial) factors. For example, INP concentrations are almost always much lower in the coastal Arctic (Wex et al., 2019, cited in the manuscript) than near the equator (for example the Cabo Verde data used in this study, Welti et al., 2018, also cited already). In the coastal Arctic, INP concentrations are also much lower in winter than in summer, and are probably even lower in the middle of the ice pack where no identified INP sources exist. Even in Cabo Verde, where the data used in this study originates, Welti et al., 2018 mention that the high time variability of INPs is partly explained by emission sources and meteorological contexts. The issue here is that the approach presented in the present study can produce very high INP concentrations where and when they should be very low, for example during the Arctic winter. It is true that the Fletcher parameterization which also be biased in conditions that deviate significantly from median conditions. However, in the parameterization of Fletcher, the bias on INP concentrations will be limited by the fact that the median INP was used to build the parameterization. In the present study, extremely high INPs can be predicted at sites where they should be extremely low, something that does not occur when using relationships like Fletcher. In low INP cases, this is made worse by the issue discussed above in Major comment 1).

3) Unless I am mistaken, the parameterization of Fletcher predicts the total number of INPs, not just immersion mode INPs (it is based on observations of primary cloud ice, when INPC=Ni the

cloud ice crystal number). While the exact details of what is included in the STD control simulation are not clear, the present study seems to replace the Fletcher parameterization in the model, or a ice=f(T) relationship, by an immersion-mode-only parameterization. How are other INP processes (deposition, contact) parameterized in the model? This could also explain why the median INP=f(T) value on Figure 1. is much lower than in Fletcher, and, if other nucleation modes are completely ignored, why the new parameterization produces less ice than STD.

4) The abstract mentions that "reasonable ice masses" are produced by the model, but no comparison with observations of the new model is provided in the paper. Section 2.2 suggests that the original model underestimated cloud ice unless unrealistic INP concentrations were used, but the updated parameterization (F22) produces even less ice, so is it not degrading the model performance?

Other comments

1) The parameterization is said to be produced from observation data at Cabo Verde and on ships, but the comparison between observations and the parameterization is never shown. Can you add observed INP data on Figure 1 or add it as a supplementary figure?
2) L.49-50: I think this is slightly misleading because some of the observed variability can actually be explained by these meteorological and aerosol parameters. Complex models do not reproduce the full observed variability of course, but that is true of any model of any quantity. In addition, as mentioned in the introduction, the models of Marcolli et al., 2007 and Wang et al., 2014 do parameterize some of this enhanced variability.
3) L. 54: There is also an issue here, because the distribution presented in the paper does sample from the global variability of INPs, but is probably worse at representing the local (spatial and temporal) variability of INPs than aerosol-aware models, which for example will predict lower INPC over very remote regions where INPs are rare, and higher INPC over deserts where dust sources are plentiful.
4) L.73-74: Isn't this in contradiction with the discussion above of how variable INPs are in time and location? These observations used to build the parameterization should also be shown alongside the parameterization or in a supplementary figure.
5) Equation(1) – Why does the formula for μ gives negative values for T>-10C? Is there a typo here? I also think all the terms should be defined with units explained, is INPC supposed to be in #/m3?
6) L. 80 What do you mean by "normalizing the distribution"? What quantity is used to normalize it and why? A lognormal distribution should also in theory already be normalized. I suppose the issue could be that equation 1 is missing a 1/INPC factor in order to be truly lognormal, unless the distribution was meant to be defined in units of dD/dlog(INPC) (since dlog(INPC)/dINPC = 1/INPC).
7) L. 82 si>0 is only explicitly required for deposition freezing. For immersion freezing, there needs to be supercooled liquid droplets present (T<0., sw>0. and liquid droplets present). In practice, when the system is in equilibrium, si will always be positive where supercooled liquid exists, but this is not directly required for immersion freezing to occur.
8) L. 87-98. How are other nucleation modes treated? Even for immersion freezing only, INPs can also cause freezing of rain so why not also take into account the concentrations snow and other frozen precipitation when subtracting from INPC?
9) L. 96-98: A new drawing from the distribution will regenerate INPs, so I think that if the drawing frequency is smaller or equal to the time step this cycling does not occur. In addition, does this

mean that the drawing frequency, when larger than the time step, is also assumed to be the INP replenishment time in the cell by e.g. advection and resuspension?

10) L. 103-111 – Instead of discussing representativity with numbers of grid points for a given hypothetical cloud, I think it would be more clear to simply present this discussion as the representativity sample size of 1000 points.
11) 2.1.1 and 2.1.2 Since the sampling of the distribution is done independently in each grid point, INP concentrations can be extremely different in a grid point and the ones directly next to it, but it is probably not the case in reality, where the temporal scale of an aerosol plume can be several hours or even days, and the spatial scale several 10s of kms or even 100s of kms (for example Saharan dust events).
12) L.128 – Is the doubling of computing time 1) for just F22 compared to Fletcher, 2) for the microphysical calculation only or 3) for the total simulation time? If 3), how can you explain such a large cost from only adding a random sampling calculation to the model?
13) L. 130 Does the standard MIMICA (STD simulation) include Fletcher nucleation or is it with constant Ni, or something else? Description of the model in 2.2 indicates that Ni can vary since it is corrected if falling above 200/m3.
14) L. 144 – Can secondary ice production in the cloud explain these ice biases, or has this been ruled out?
15) L. 162 – How was graupel and snow removed and can clouds precipitate or be removed by another process? Can this removal cause issues, for example, if there is no precipitation in the model have you checked that the cloud droplets and ice crystals do not grow to unrealistic sizes?
16) 2.2 and l. 171-172 – Can you explain more explicitly what are the differences between STD and F22, is this just the ice nucleation or are there other changes to how cloud ice is handled? Is graupel and snow also removed from STD?
17) 2.2 and Figure 3. - You mention that the issue in STD is the difficulty in reproducing observed cloud ice, but the IWP is even lower in F22, is this not degrading model results?
18) L. 181 - Can you provide a number estimate or figure showing this latent heat or temperature increase in the model?
19) L. 183- Can you also plot the water vapor column to show this gas-phase increase? The water budget should be balanced and it should be possible to show this shifting of water around phases directly.
20) L. 185 – Unfortunately, this increase with time is probably at least partly caused by the issues with the parameterization discussed above, because with enough time, INPC values large enough to freeze any supercooled droplet concentration will be sampled from the distribution. It seems that when discussing the sensitivity to drawing time the slope of the increase is indeed larger for smaller drawing times and the gap between experiments grows during the simulation.
21) L. 186 Can you also calculate here directly this effective cloud ice crystal mass from model results of ice crystal numbers and ice mass concentrations?
22) L. 231- 232 - This is very interesting and probably true for other parameterizations. For this reason, a parameterization only based on a median of INP observations will probably underestimate primary ice formation and some sort of correction might be necessary to consider short term variability and the likelihood of efficient INPs deviating from the median values. However, I am not sure of how applicable to other approaches this is. For example, earlier simple parameterizations like the one from Fletcher were I think directly based on observed ice crystal numbers, and did not suffer from this issue directly. And I think that some of the more complex aerosol-aware parameterizations will assume that these efficient INPs are quickly removed from the atmosphere and will not continue to nucleate ice later.

23) L. 260-261 – Is there a quantitative estimate of INPC variability at very short time scales (a few minutes and meters) in these references? INP are often observed at very low resolutions (very few observation sites, daily resolution or lower) so I am not sure how large this small-scale variability is, compared to the large-scale variability used to build the parameterization.

References:

Natalie Matosin, Elisabeth Frank, Martin Engel, Jeremy S. Lum, Kelly A. Newell; Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. Dis Model Mech 1 February 2014; 7 (2): 171–173. doi: https://doi.org/10.1242/dmm.015123