Response to Reviewer #2

The authors appreciate the through reviewing and would like to thank the reviewer for taking the time and for helping to improve our manuscript, especially it's qualitative nature.

This work attempts to build on the extensive past literature on droplet freezing assays, by leveraging microfluidic technologies. Several calibrations experiments are performed on the stage in an attempt to quantify the accuracy and precision of measurements of ice nucleation rates and ice active site densities. It is noteworthy that assessing the absolute values of these quantities is a particularly challenging endeavor, owing to the lack of 'primary standards' whose nucleation behaviors are known to a high degree of certainty. Indeed, past intercomparing studies have yielded considerable spreads in rates and ice active site densities for homogeneous nucleation and the heterogeneous nucleators, and deviations between instruments has not been satisfactorily addressed. While the current work could definitely be of interest to the community of researchers interested in atmospheric ice nucleation processes, there are serious gaps in the paper which I feel need to be addressed prior to consideration for publication. The majority of these gaps revolve around the qualitative nature of the comparisons performed, and lack of error analysis on the rates and ice active site densities determined.

Major issues

- In figures 5, 7, 8 & 9, there are no vertical error bars presented to represent the uncertainty in the measurements. What is for example, the uncertainty stemming from the stochastic nature of nucleation in the homogeneous freezing experiments? What is the effects of the uncertainty stemming from the random sampling of ice nuclei into droplets on the heterogeneous ice active site densities in figures 7, 8, 9. Without statistically sound error bars and confidence intervals on the certainty of the data, comparisons are rendered almost meaningless.

Authors' Reply: In figure 5 error bars are added for WISDOM and for the rest of the data that is used. In figure 7 the errors were confined in the marker itself, both for x and y axis. We preferred different marker sizes because visually it was hard to place the error bar in a clear way. This is now mention in the figure caption:

"Accumulated active site density spectra (n_s) of K-feldspar and Illite-NX particles as a function of temperature from validation experiments of immersion freezing in WISDOM. Frozen fraction values are represented by a color bar, for few surface area values that are exposed in 40 and 100 μm droplets. The dependence of the nucleation site density on the surface area is illustrated here. WISDOM uncertainties are included within the size of the markers. The uncertainty in n_s is propagated from the error in the surface area and the error in the frozen fraction. For validation, previous immersion freezing measurements are also presented (Hiranuma et al. (2015) and Atkinson et al. (2013)). For the Hiranuma et. al. fit, the maximum deviation between maxima and minima in the vertical axis are shown by the error bars for the relevant temperature range."

Moreover, new graphs with the error bars for the illite and for the k-feldspar data is now added.

Authors' Reply: In figure 8 and 9 error bars are added.

- The authors note in the abstract (L16), that the method produces excellent 'statistics'. To what quantity are the authors referring to here? Nucleation rates? Ice active densities? If so, what is the effects of sampling hundreds of droplets over say 50? By what kind of factors are uncertainties reduced? If this is purported to be a major advantage of the technique, surely the improvement in these 'statistics' by this method should be quantified?

Authors' Reply: Due to the concern of both reviewers regarding the term "excellent statistics", this was removed from the manuscript:

"Frozen fraction, ice nucleation active surface site (INAS) densities and freezing kinetics can be obtained from WISDOM measurements using hundreds of individual droplets in a single freezing experiment. Extensive calibration experiments using eutectic solutions and previously studied materials are described".

- in this direction, some of the benefits of this technique over past techniques should be elaborated upon and clarified; the discussion of past issues, without acknowledging how they have been successfully dealt with in the past is rather peculiar. For instance, many cold stage instruments do not suffer from issues surrounding the Bergeron-Findeison process, and do not need oil to be placed on the droplets. At line 53, it is said that generation of 1 uL volumes is not trivial. To my understanding, this can be done with a pipette, which would seem rather trivial to me.

Authors' Reply: In the manuscript we point out in section 3.6 the advantages and disadvantages of the WISDOM and provide our considerations for choosing the microfluidics approach. We further show the importance of a thorough calibration of temperature and temperature equilibration properties.

Clarification: in line 53, it is claimed that smaller volumes are not trivial to generate and not 1 microliter droplets, which is indeed trivial.

- In section 2.3, on the automated detection of phase transitions, it is noted that the algorithm can 'successfully distinguish between a phase transition event and noise' (L 150). Whilst this statement may well be correct, I see no mention of to what accuracy the algorithm can successfully distinguish between phase transitions. Is this 100% accuracy? How many experiments were performed manually to determine this?

Authors' Reply: This is monitored for each freezing event and done separately for each experiment since the noise level varies between experiments. The freezing causes to larger reduction of the grey level than the noise, and hence it is possible to separate it from the noise. So a threshold is being set to differentiate the noise. This leads to very high success rate in identifying the freezing events.

• In the quoted value of ±0.25 K for the Linkam cryostage temperature sensor, which is subsequently quoted in the captions for figures such as 5 and 7, how was this value determined? Knowing this would certainly be useful for the reader. By what procedure was this value obtained?

Authors' Reply: This value (<±0.25K) is the uncertainty of the Pt100 temperature sensor in the temperature range. Using this value with our calibration results we propagate the total temperature uncertainty (±0.34K). This was not clear in the text and thank you for this comment it is now explained in the text.

• In figure 7, the data for NX illite appear to be at the extreme lower end of the spread, based on the error bars used for the Hiranuma data. Surely this should be discussed in the text?

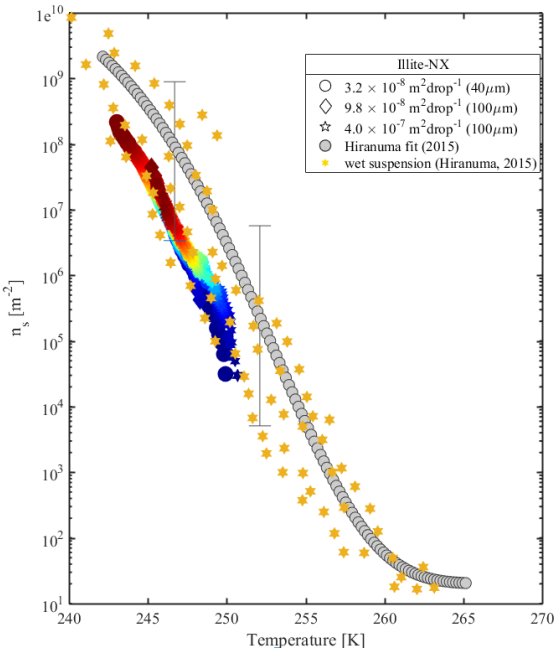
Authors' Reply: The comprehensive data of Hiranuma et al. covers different instruments and several of immersion freezing techniques. Hiranuma et al. also mention the spread of the results, up to three orders of magnitude. In the temperatures range where WISDOM has freezing signal that is representative, the fit of the suspension groups contains only two instruments (CU-RMCS and NC-State-CS) that are relatively spread from each other, in comparison to the spread among most of the other instruments. WISDOM and CU-RMCS agree well within the same order of magnitude on the number of active sites. Investigating in details the reasons for the spread in the results of all instrumentations is beyond the scope of this paper and discussed in Hiranuma et al.

• In lines 268-269, it is said that ns is in best agreement with the Leeds-NIPI for NX-illite. Yet close inspection (the subtle shades of grey used here in the graph make this a bit difficult to see), shows that there is in fact no overlap in the temperature range between the measurements presented here, and those of the Leeds NIPI μL.

Authors' Reply: Explanation is united with the reply to the next comment.

• In addition, in comparison to the binary instrument in figure 7, the data are up to an order of magnitude or greater off, which is not immediately obvious as the authors have chosen to only label the scale for every factor of 100 increase. It is noted that this is within the uncertainty of the instruments, but what is the uncertainty of the quoted values for the WISDOM (see my first point above…). Does the uncertainty really cover 2 orders of magnitude? With what degree of statistical certainty are you sure that these two measurements are in agreement?

Authors' Reply: The conclusion that the Illite-NX data is in good agreement with the BINARY rely on the fact that there is less than one order of magnitude between the two data sets and this difference is justified by uncertainties of both instruments. The temperature at which there is more than an order of magnitude difference or more (~250K) is where the variation between our measurements are the highest and for < 0.01% of the droplets. This is where rarer active sites exist (in the used surface area) and is less representative of the material. Moreover, if the slope of the Leeds NIPI is extrapolated, it would nicely overlap with WISDOM data. Here it is also expected that the data is of higher significance as there are many more freezing events. Examination of the Hiranuma et al data and our data, it seems that the data converge to the WISDOM and CU-RMCS results and slope:



*The labeling of the y axis is changed to every factor of 10 increase to avoid confusion of the readers.

• Lines 136-137: If the chips are being clogged by larger particles, then you may be severely altering the size-dependent particle composition of the samples as they pass through. How is this dealt with and accounted for?

Authors' Reply: This is indeed a concern and we do not work with clogged devices d or with materials that tend to clog the devices. Moreover, we repeat experiments with the same device for few times (with fresh droplets), and do not observe evidence of critical alteration in the size distribution. We would expect that such alterations will increase with increasing number of repetitions and that we will see a reduction in the ice formation efficiency. For materials with particles >5 micron in diameter, we experienced immediate clogging. From our experience, these particles readily settle in the suspension, even before the droplet production and also in cases where it was stirred during the droplets production. We also expect that for materials that tend to clog the device, it will be difficult to quantify that as it might change with the flow used or concentration of the material in the suspension and the material's size distributions. Hence we do not work with suspensions of larger particles or when we observe evidence for clogging.

Other issues

• Line 28: INP should be INPs

INP changed in the text to INPS. Thank you!

• Line 34: Why is this only 'possibly' in future climates?

How can we be certain that this will affect climate?