Referee comment by Gabor Vali on the "Fifth International Workshop .... " by DeMott and co-authors.

This manuscript reports the results of a workshop held in 2015 at the Storm Peak laboratory in Colorado with a large collection of ice nucleation and aerosol instruments. The main goals were to assess the degree of agreement among ice nucleation instruments of various designs and to interpret the results with consideration of aerosol characteristics. These goals have been mostly achieved. The agreements found are significant and the authors also make clear where discrepancies appeared. The results are described in detail.

Since the process of ice nucleation is complex, inferences about how it takes place in the atmosphere or in the laboratory have been arrived at over many decades from the way the process is manifested in instruments that proceed from prior condition to the appearance of ice with controlled characteristics. There are inherent limitations to this approach as knowledge is hard to achieve of relevant aerosol properties and of the way conditions are altered on the way till observation of ice becomes possible. Yet, by gradual understanding of the impacts of instrument design approaches have been key to the gradual development of a degree of understanding of how to achieve usable measurements of heterogeneous ice nucleation by atmospheric aerosol particles and by some types generated from known substances.

The workshop described in this paper undertook to test how instruments which have been shown to yield agreements among themselves with controlled aerosols will perform in the relatively clean high-altitude environment sampling ambient aerosol. Similar efforts have been reported previously, but this paper reports on significantly more extensive measurements. No surprises emerged from taking the devices to high elevations and, perhaps, that can be taken as reassurance of their reliability.

The paper is long and demanding. Instruments and sampling setups are detailed, possibly to a greater extent than necessary since previous publications are available for all. On the other hand, the results are presented in a more compressed style. The aerosol data are used to interpret outliers in the intercomparisons but this is somewhat cursory. Going beyond the intercomparisons, testing agreements with assumed particle compositions and parameterizations (Section 3.5) could well have been a separate paper. That section and the aerosol data needed for it take up, by a rough estimate, about 1/3 of the paper. Similarly, the deposition mode data, from only one instrument, could be separated from the main subject.

There is little attention in the paper on the absolute values of INP concentrations observed. The mountain-top observatory could be expected to yield INP concentrations lower than those observed a low altitudes. This doesn't appear to be the case.

The main result of the intercomparisons of instruments is stated as an order of magnitude agreement for immersion freezing INPs. This is a good result from the perspective of instrument reliability compared to, say, what the situation was 20 years ago. On the other hand it is worth seeing it as a 5ºC discrepancy for given INP concentration. For atmospheric cloud processes that is still a huge uncertainty. Thus, for INP measurements to be useful predictors of cloud evolution at given location and times there is much remains to be accomplished. This does not diminish the accomplishments demonstrated in the paper, but it may be worth reflecting on in the Abstract and in the Conclusions.

On its own terms, the evaluation of prior parameterizations with the extensive aerosol data available from FIN-03 is a laudable idea. Combined with similar efforts a better appreciation is developed for the validity of the generally used predictive equations for atmospheric INP concentrations.

Essentially all data reported is for measurement temperatures of -15°C and lower. This limitation should probably made clear to the readers right up front. The reasons for this limitation also deserve to be stated. A

further caveat applies to having all comparisons done in terms of the cumulative number concentration, with only brief attention to the slopes of the temperature spectra.

The degree of agreement among instruments is extensively analyzed with reference to Figs. 6 and 7. Another striking point that can be added to that analysis that while FRIDGE-CS and CSU-IS show significant changes with time between Sept 14 and 16, specially at the higher temperatures, these changes are not evident in the other data. Then, for the next two days, these two instrument show greater disagreement in Fig. 7.

In the title, placing importance on what was done, rather than on the vehicle for doing it would be useful, i.e. "Field Intercomparison of Ice Nucleation Measurements; the Fifth ....."

Line numbers:

77-176 The second part of the Introduction is not as clear as it should be for readers not already familiar with the subject. More specific comments are given in the following.

120 The text of the preceding lines calls for a somewhat more detailed reference to history. Ice nucleation comparison workshops go back as far as 1967. The main reason for many repetitions is not a difference in goals but the development of new instruments and better characterization of the processes underlying the different instruments, i.e. better understanding of heterogeneous ice nucleation. Another important factor was growing recognition of having detailed aerosol characterization (lines 169-172) and better instruments for that information. All of these factors see advances, but important limitations also remain.

121-150 Following on the previous comment, it would be helpful to readers to have a concise argument for why FIN-01 had to be followed by -02 and 03. Why was -03 deemed necessary, The contrast with Knopf et al (2021) and Lacher et al (2024) would be best detailed at the end of characterizing FIN-03 goals.

210-215 This may not be as simple as it appears here. Not all sizes of aerosol are of the same composition even at a given instant of time and the INPs are a very small subset of the total aerosol, so the manner in which aerosol composition is correlated with the INP measurements needs detailed description and substantial caveats. The subject should perhaps be addressed after the description of the instrumentation.

235 " ... positive or negative mass spectra" - what does that mean?

236-241 This helps to clarify what was meant on lines 210-215, but only in part. Are the sample sizes of the two instruments the same?? If PALMS samples only a fraction of the particles sampled by LAS, how much uncertainty enters in the analyses?

Section 2.1 The authors' approach to present how data from given instruments are interpreted along with the description of the instrument has merits. Caveats and limitations are stated up front. However, it separates those issues (some listed in the foregoing) from the actual interpretation of the results. In the end, I think it isn't working to the paper's advantage.

313 "All these instruments ..." refers to the types of devices or the specific units?

326 Please clarify " ..substantial temporal overlap ..". Probably not easy to express in a few words but it does reflect importantly on the results, specially if the temporal variation of aerosols is considered. A graphical representation of the sampling timeline along with, say, total aerosol concentration may be useful.

334-335 'several prior works' and 'several publications' are unnecessary words

379-382 Bit unclear. Seems to indicate that tests were made during FIN-03 and the factor 90 increase found previously was confirmed.

413-417 The vagueness of this description leads to unease about the results. What is meant by 'non-ideal' behaviour?

418 "... was then applied ..." When?

422 it is unclear what "linear interpolation" refers to. Between particle-free sampling periods? How often was that done?

425 " points that exceeded water saturation" seems to be an error

439 Sentence appears garbled and it is unclear what 'complementary' means here.

442 ".. distributed liquid particle suspensions .." - unclear

466 This is the first mention of the fact that off-line measurements were also off site. Please include that information at some more prominent point.

470 ".. detected by an optical microscope.." - needs better wording

472 The dependence on cooling rate is known and can be corrected for variations. If the other instruments used a different rate, the comparison can be improved by applying the known corrections.

478 Dilution has not been mentioned before.

479-480 More accurately, spectra are calculated for $f_{unfrozen}$ determined at 1ºC intervals. Not so?

482 Were blanks processed for NC State CS?

508 " ...special contribution ..." One wonders why this is considered special.

508-540 This material reads more part of an interpretation of results than a methods dscription.

547 Presumably, 'standard method' refers to measurement with this device, not in a more general sense.

562 Is the EAC used in FIN-03 the same as the PEAC7 described in the reference? More importantly, are the size-dependent collection efficiencies of this EAC known and were corrections made to account for it. This is specially relevant to INPs potentially of sub-micron sizes.

586 " ... attempted to operate ..." leads to questions

588 Presumably, "agreed upon" for reasons of surpassing detection limits and to get statistically reliable results.

582 Section 2.3 seems to have material that should be elsewhere. The sampling strategy should probably be clarified earlier on. See comment on line 316. The range of temperatures for inter-comparison is a matter of detection limits and, to some extent, a compromise with considerations of sampling times.

725-740 This brief reference to the slopes of the temperature spectra is, apparently, based on some visual fit of the curves in Fig. 5. More could be said about how this was done and much more should be said about variations in the slope with temperature and with time. The slopes at given temperatures show significant differences among instruments.

726 Is the MIT-SPIN data for a single measurement or average. If the latter, why no error range is indicated? This sentence should be in the preceding paragraph, not when the temperature trend of the data is discussed.

733 Unclear what is meant by ".. standard deviation of the measurement means ..". Is it mean values for the 1ºC interval? Once the data has been set at 1ºC bins, as stated earlier, they do not need to be called 'mean' values. It is also a bit unclear that the mean and standard deviation refer to averaging all values at a given temperature throughout the campaign. Assuming that, it would be helpful if here, or at some other place, the number of measurements represented by each point would be given.

738-740 Any suggestion for explaining the difference?

743-745 Can the difference be quantified and its significance assessed in relation to data scatter?

747 Wasn't the FRIDGE sampling inside the laboratory, while the other two on the roof?

751 ... Except for the online instrument, what is meant by average over 1ºC intervals? Why these 3-hour averages differ from the 4 hour sampling times?

768 Figure 6 is probably the most complete presentation of the intercomparison results, yet it is hard to extract much information from it. The vertical scales are very compressed, and too many lines are presented. Could fewer temperatures be included? The color code is overly subtle, perhaps lines for -20ºC, and -25ºC could be heavier to help orientation.

772 See question about 3-hour average for line 751.

827 " ... higher nucleation efficiency .." might better expressed as having higher likelihood of containing ice nucleation sites..

848-850 Breakup in itself is not an obvious explanation for reduced ice nucleation activity, unless sites are assumed to be formed at component contact areas. Various other factors also come into play when particles are introduced into water.

865-870 This paragraph deserves a second look. First of all, error sources may be un-assessed but not ultimately unquantifiable. Second, cooling rate dependence is pretty well quantifiable for cold-stage devices. The effect of sample storage has been assessed as referenced on line 868. Instrument temperature drift is in contradiction with the stated accuracies. Filter and impactor sampling has substantial possible uncertainties for unknown INP sizes. Many other sample handling issues may be variable to some extent.

880-884 Do the p-values here refer to no discernible offset?

899 What does IS stand for?

1186 A bit of a stretch to indicate that the range of measurements extended to -5ºC. Comparisons were all for -15ºC and lower (Fig. 6)

1201 Closure may refer here to the aerosol/INP connection. INP/cloud-ice connection is another, so this distinction should be made. Regarding the aerosol/INP closure, there is a limit beyond the instrumentation complex here utilized in that INPs may constitute a subset of the aerosol different in composition and size than the predominant aerosol.

1256 The choice of wording ".. relative agreement .. " makes one wonder as to relative to what. Perhaps 'the degree of agreement' is less vague as it can be thought to refer to the degree established in the analyses.

1181 The authors should consider including in the Summary some comments to weigh the atmospheric relevance of the results with regard to the temperature range covered by the measurements and the uncertainties of the relation between INPs and ice development in clouds.