Response to Gabor Vali's comments on "The Fifth International Workshop .... " by DeMott and Co-authors

We thank Dr. Vali for his comprehensive and insightful comments, and his understanding of the complexities of such workshops.

Broad comment [1]: While it is clear that no classification can reflect perfectly the variety of instruments used, it is worth thinking about how useful is the one applied in this paper. No doubt, the online vs. offline designation has its origin in common references to the instruments in that some can be deployed in the field as a self-standing unit, while the others have essential work done in the laboratory. Thus, 'field' vs. 'laboratory' instruments could also be applied. The current designation has disadvantages. It doesn't inform people unfamiliar with the details of what the essential difference is. Is it the operating principle of the devices? Is it aerosol flow through the device versus samples captured and processed? Portable versus fixed in not necessarily applicable, as a filter processing unit is no larger and difficult to transport as a CFD chamber. A drop freezer can be the smallest of all the devices. Also, field deployment of a device impacting a flow of aerosol on water drops would be an online instrument for which much of the RH discussion about CFD's would be irrelevant. So, while the offline/online distinction isn't all bad, perhaps "direct sampling" versus "post-processing" would be more meaningful. Another useful direction would be to focus on variable RH instruments versus those with liquid water. What characteristics is the most important? Tables 1 and 2 do a pretty good job in listing the main features of the various devices. But the titles of subsections 2.1 and 2.2, and the headings in the Tables could be improved by removing the hesitation in the designations and, preferably before 2.1, explain the reason for the classification. Dry versus wet in the table headings is not the best. My recommendation is to use 'direct sampling' versus 'post-processing' and define in the Introduction how these terms are applied. This is also appears to be the intention with lines 30-32 on page 7. These designations would allow the AIDA chamber to be in the group of instruments with clouds forming on the sampled aerosol, just like in the CFD's. Omit 'portable' as part of the definition. (Apologies for this rambling paragraph.)

Authors' response to Broad Comment [1]:
It was a most difficult decision, in considering how to separate the discussion pre-submission. However, the workshop was separated in a distinct manner that the discussion mimics. The samplers for offline immersion freezing measurements required the largest particle loadings and so sampled first each day. This sometimes included two separate fills of the APC. At the end of this period, online instruments would collect particles for a period from the APC, over a period that extended a couple of hours beyond the period of sampling for post-processing. Finally, an aerosol fill was done directly into the AIDA chamber and this accommodated the few of the post-processing samples for devices that did not desire quite as high particle loadings (e.g., FRIDGE-STD and DFPC), and then other online samplers that would sample from AIDA just prior to the AIDA expansion. The concentrations were never the same in the APC period and in AIDA. Importantly, the aerosol fill to the AIDA chamber was not always the same INP type used in the morning sampling period from the APC. This was a conscious (and deemed necessary) decision to accommodate the ability of all instruments to sample all aerosols, and for the direct sampling

instruments to be able to do so over a range of temperatures (e.g., only limited temperatures could be assessed by flow chambers in the two hours after the collections from the APC into liquid or onto filters were done, and before the AIDA expansion had to begin). We mention this here to preface our responses to later comments. Nevertheless, understanding this may help to explain a lot about how the paper is organized, and hence we have decided to add such discussion to the Methods section. We hope that this discussion helps all readers to understand why we use concentration comparisons in some plots and then revert to active fraction and/or active site density in later figures that include data from both APC and AIDA experiments.

As for the exact description of device types, field versus laboratory would not work, because some of the immersion freezing methods are being taken to field facilities for immediate processing, mitigating in some cases the need for freezing samples. Direct sampling versus post-processing is more or less what we intended as the meaning for online versus offline, so we accept Dr. Vali's suggestion on using this terminology.

Changes in manuscript re: Broad Comment [1]:

This discussion will appear as follows as the 4th paragraph in revised Section 2.3, with the new title, "*Generation of varied INP types and general study procedures*". We write, "*The daily protocol determined for aerosol generation and measurements is an aspect of these studies that bears strongly on the organization and discussion of results in this manuscript. Especially, sampling periods were organized to optimize the opportunities and conditions for all instruments to sample the variety of aerosols. Each day over the three-week workshop period typically began with fills of one INP type into the APC, and sampling of that aerosol into liquid and onto filters over a two-hour period for later assessment by the post-processing devices. This was followed by the direct samplers processing the same INP type from the APC over another two-hour period. This typically permitted measurement at a couple of specific temperatures, with data at other temperatures being acquired on another day for the same INP type (from the APC or AIDA). Then, at midday, the AIDA chamber would be filled, typically with a different INP type than used in the morning APC experiments. Direct sampling from AIDA by the flow chambers would occur over a period of time just prior to the start of expansion cooling experiments. As well, collections onto filters or wafers used by the DFPC and FRIDGE device (standard method) would be made only from the AIDA aerosol fill on each day, since these methods required very short sampling times in order to limit particle loading that could impact vapor depletion in these diffusion chambers that employ particles on substrates, as already discussed. For example, the DFPC-ISAC filters were collected periods of 10's of seconds. Other collections into liquid or onto filters for immersion freezing post-processing would not occur from the AIDA chamber prior to expansions. A consequence of these procedures is that we will find it convenient to present results on different bases when discussing sampling from the APC and from AIDA. While we might ideally wish to present all data on the same basis as measurements are reported for atmospheric sampling, as number concentrations, we choose to do so only for the APC experiment period that offers the opportunity for comparing the most measurement systems. We describe how that is done next. For AIDA sampling, we will display results as active fractions and ice active site density, which then allows integrating APC results along with AIDA results for the same INP types over the course of the workshop.*"

Additionally, we widely employ the terminology of direct sampling and post-processing for the different methods used. We introduce this in Section 2 with the statement, "*These categories are, firstly, instruments operating online or for direct processing of aerosol particles and, secondly, those utilizing collections of particles for subsequent offline or post-processing*." We also revise section headings, table titles, and consolidate the AIDA chamber description into the direct sampling category discussion in Section 2.1 and remove the original Section 2.3.

Broad comment [2]: I am somewhat reluctant to raise this issue, but this paper needs special scrutiny in this regard because it weighs heavily on the principal results of the paper. Screening of data to be included or excluded is perhaps the trickiest and most sensitive part of the paper. There is a sense of some censoring of the data throughout Section 2. Undoubtedly some judgements had to be made about data validity by some criteria. One can see the attempt to do this, and the authors certainly deserve to be given the credit of proper judgement. However, the first question that comes to mind when looking at the degree of agreement shown in the results is how much screening of the data was done? How many discrepant cases were excluded due to errors or uncertainties? In other words, how much subjectivity entered into the analysis. Some comments on this issue would help the readers gauge the value of the results.

Authors' response to Broad Comment [2]:

It is disconcerting to hear the word censoring raised in regard to judgements on instrument operation and data submission. While full participation was encouraged, even for groups doing such a thing for the first time, there was no requirement for individual groups to submit data for every case covered in this paper. In fact, there was no requirement to collect data in every case, and it is the case that some groups collected much more data than others. We have tried to be upfront about the fact that this workshop had both formal and informal parts, and this paper deals with the latter. Investigators who recognized a certain issue in a given experiment that was not part of the formal intercomparison (a separate paper) were free to remove the data pending their own investigations (some that are continuing). This was not a frequent occurrence. In at least one case, a temperature calibration issue was recognized by one team (M-AL) following the workshop, and all of their data were subsequently reanalyzed. The MIT SPIN group did not always extend RH scans to levels sufficient to express immersion freezing in all experiments during the informal workshop phase. Other groups chose to focus their efforts on the formal period of comparisons, and to use the other part of the workshop for development activities (PNNL-CIC). We do not wish to state all of these details in the paper. It should be clear by the volume of data shown in many of the figures that this allowance was not in any way an attempt to unify the data sets. The participation of different groups at different levels can also be recognized in the data table listing. We considered that if 15 measurements are used to demonstrate correspondence or not, that this is a reasonably sufficient examination.

Changes in manuscript re: Broad Comment [2]:

We re-emphasize at the end of the Introduction that this paper describes the informal component of FIN-02. We write, "*This paper is intended to overview the informal activities of the workshop while addressing the majority of*

these objectives. It is not intended to answer all of the goals and objectives that are better addressed in separate studies. It is not our intent to rigorously test the capabilities of different measurement systems, but rather to point to areas of success and areas for needed development or further research. For these reasons, and to include as many measurements as possible in comparisons, we focus primarily on measurements relevant to immersion freezing nucleation, as discussed further below. This allows for integrating the most possible measurements into comparisons made for assessing one important aspect of the state of the art of ice nucleation measurements." In the Conclusions we write, "*Some operational issues occurred for investigators at times (obvious errors, measurement biases, inability to achieve comparative conditions for proximal immersion freezing) and where these were recognized, data were either not entered into comparisons or required revision. Some issues were investigated, such as the appearance of small ice in the CSU CFDC data for INPs with steep activation functions. Others remain the subject of active investigation."*

Broad comment [3]: It becomes gradually clear as one reads on in the paper, that the focus of the analyses was freezing nucleation, at least as much as it could be assumed to be the dominant mode of nucleation if sufficient supersaturation was achieved to ensure that INPs are in liquid droplets. Perhaps, I missed it, but if this is true, shouldn't it be clearly spelled out in the Introduction, and be part of the goals? Similarly, the size range covered (9/28-32) is a pre-defined constraint and would be best stated up front.

Authors' response to Broad Comment [3]:

We are thankful for this comment because it was not intended that this be a gradual realization. In the abstract, we had stated the different measurements with a focus on immersion freezing (modified here to account for the revised naming of the different methods: "The results presented here use data from the workshop to assess the comparability of immersion freezing measurement methods activating INPs in bulk suspensions, methods that activate INPs in condensation and/or immersion freezing modes as single particles on a substrate, continuous flow diffusion chambers (CFDCs) directly sampling and processing particles well above water saturation to maximize immersion and subsequent freezing of aerosol particles, and expansion cloud chamber simulations in which liquid cloud droplets were first activated on aerosol particles prior to freezing." In the first paragraph of the Methods section we state (again, slightly revised here in new form), "No measurements of contact freezing were included in this study, and neither will we discuss results herein from workshop measurements that were made in the regime associated with deposition nucleation (including at temperatures below the homogeneous freezing temperature or pure water droplets), but we will focus on inter-comparisons of particles acting via immersion freezing or proximal behaviors. By proximal behaviors, we follow the terminology of Vali et al. (2015),…" We simply could not pull all of the workshop studies into this overview paper, or it could have become unwieldy. We attempted to point out that the database remains for future papers by the participants, or inspection by any interested parties.

Changes in manuscript re: Broad Comment [3]:

In addition to the modest changes mentioned in our response above, we now add a direct statement at the end of the Introduction pointing out the emphasis decided for comparisons presented in this paper.
"This paper is intended to overview the workshop while addressing the majority of these objectives. It is not intended to answer all of the goals and objectives that are better addressed in separate studies. For that reason, we focus here primarily on the immersion freezing process, as discussed further below. This allows for integrating the most measurements into comparisons made for assessing the state of the art of INP measurement systems."

Broad comment [4]: Unless I missed it, the promise made at 8/32 to describe the aerosol collection and transfer methods for testing as suspensions are missing from Section 2.4. Problems associated with the method used and the final efficiency of the capture are not found. See also 13/4-5.

Authors' response to Broad Comment [4]:
We thank Dr. Vali for pointing out that we indeed neglected to describe the details of particle collection into liquid.

Changes in manuscript re: Broad Comment [4]:
We have added this now to Section 2.3 (formerly 2.4). The segment below has been integrated into a revision of the overall discussion of experimental protocol for APC sampling. In doing so, some repetitive statements are now removed.
"Collection of particles into liquid suspensions for shared use by a suite of immersion freezing devices was performed by impinging a flow of particles from the APC into a glass bioaerosol sampler (SKC Inc.) (Hader et al., 2014; DeMott et al., 2017), referred to here as impinger samples. Two impinger samples were collected for ~120 min with a flow rate of 12.5 L min^{-1} from the APC. Flows were checked daily. Impingers were cleaned by wiping, rinsing with ultrapure water (18 MΩ cm^{-1}), and soaking in isopropyl alcohol overnight (2-propanol, ≥99.8%, ROTH). Before assembly the impingers were rinsed using ultrapure water water once more. Following the sampling of Snomax® particles, the impingers were baked overnight at 200°C instead of soaking in alcohol. This was done to eliminate the possibility of carryover by ice active due to these biological samples. During sampling the water was replenished every ~30 minutes to keep the water level near 20 mL. Due to evaporation, the final bottled volume was typically about half of the added water. The two impinger suspensions were combined into one sample and topped off to a total of ~36 mL. The sample was divided into 4 ml aliquots, bottled in pre-cleaned DNA free cryovials and stored locally in a freezer at -20°C. The same procedure was applied to the handling of blanks. At the end of the campaign, blanks and samples were placed on dry ice and shipped overnight to participating groups. Shipment to Israel was delayed by customs, allowing the sample to thaw en-route. After receiving the sample, each group decided on their own sample storage and handling strategies. Again, we note that the University of Leeds group performed NIPI measurements of these suspensions on-site in Karlsruhe immediately after collection (i.e., without freezing)."

Broad comment [5]: Some re-phrasing of conclusions such as in lines 6 to 23 on page 13 should be considered. The degree of agreement in the data is remarkable and deserves to be celebrated. But to say that the agreement extends over 7 orders of magnitude over the whole temperature range is misleading. Except in the center of the temperature range only 4-5 data sets are represented. Thus, in a strict sense, the agreement among all data sets is not quantitative over the whole range. They all coincide in defining a common trend and agree in overlap regions. None of the data sets extend over the whole range. Also, there is a division of the types of instruments for which data are available at the upper and lower ends of the temperature regime. The main novelty in these results is the agreement between CFD type and drop freezing types of measurements in the overlap region. This is a real accomplishment and demonstrates that the two types of measurements can be combined to yield atmospheric measurements over a wide temperature range. The extent to which that is a practical solution for low INP air samples is worth addressing.

Authors' response to Broad Comment [5]:
We have revised the statements accordingly, since indeed, the point was to show the meshing of results for this particular sample, as well as correspondence overall within groupings. We remove the 7 order of magnitude statement for this reason.

Changes in manuscript re: Broad Comment [5]:
The paragraph now reads: "*The comparisons obtained for sampling Argentinian soil dust particles (Fig. 4) were among the best in this study. A striking feature of these results is the general correspondence between all methods and sampling types in ranges of overlap, as well as the apparent meshing of results from direct sampling and post-processing of immersion freezing to capture 7 or more orders of magnitude of INP activity in the temperature regime from -5 to -35°C. Direct overlap showing correspondence of the continuous flow methods with a minimum of four different bulk methods occurs over 3 orders of magnitude range at temperatures from -20 to -30°C. Good consistency is also seen amongst direct sampling methods as a group and post-processing methods taken together as a group, for shared impinger samples, and whether post-processed samples were collected by impinger or separate polycarbonate filters (IS and FRIDGE-IMM) that were subsequently rinsed of particles. Recall that the IS filter was collected simultaneously with the impinger sample, while the FRIDGE-IMM filter was collected over a shorter time frame. The largest discrepancies, in consideration of measurement uncertainties (see Supplement for explanation of measurement uncertainties for each device) occur at the coldest temperatures. In this region, data from the PIMCA-PINC instrument, which activates individually-grown droplets on particles prior to cooling, falls at the upper end of measured INP concentrations in comparison to the few other immersion methods that extended to the heterogeneous ice nucleation limit, just warmer than homogeneous freezing temperatures. We note again here that some scatter may occur in the online methods due to investigators deciding in each case what RH_w above 100% to report data for as representing immersion freezing.*"

Additionally, we have added a discussion point in the Conclusions on the issue of low INP sampling. This was mentioned in regard to the likely need to pre-concentration INPs for direct processing at >-20°C. This is also a prime

topic of FIN-03 and its overview paper. We will not add the Conclusion revision here, as the entire new Conclusions are appended to the response to Reviewer 2.

Broad comment [6]: Only in S.2.1 is the method for calculations of the INP concentrations stated. How were the presented data derived for each device should be part of the description for each one. This also applies to how uncertainty ranges were determined.

Authors' response to Broad Comment [6]:
This comment was retracted in an updated note by the reviewer. As noted, this information is available in the extensive Supplemental material, repeating in most cases information that is already published.

Changes in manuscript re: Broad Comment [6]: None.

Broad comment [7]: A suggestion I make reluctantly, because of the amount of re-writing it implies, the authors should consider including a Discussion section, moving some material from Results there, and going a little further in some themes. Having a Discussion section would allow for better perspective on the various aspects of the results (e.g. 14/3-22), and give good opportunities for comparisons with previous publications (e.g. 18/2-28).

Authors' response to Broad Comment [7]:
We can see how this might work, and appreciate the suggestion. Nevertheless, we also believe that the flow of the paper, and the manner that the AIDA studies are used to continue the comparison of data and to then fold in APC results, favor the present structure. With the introduction of plots in n_s space, it is natural to add the parameterizations and discuss them in place. It would take major rewriting to extract discussion and have it stand alone in an organized manner. Expanded discussion may not have an advantage over a more concise description of results and reference to previous studies.

Changes in manuscript re: Broad Comment [7]: We have renamed the section as "Results and Discussion" and have chosen to expand the Summary and Conclusions.

Broad Comment [8]: Parts of the description of results refer to the sigmoid shapes and the central slopes of the INP vs. T curves. I mention here that similar considerations, and a definition for a slope parameter were given in Vali (ACP 14, 5171- 5194, 2014.). The rich material in this paper could very usefully add and expand on the analyses in that paper. That may go beyond the scope the authors set for this paper, but in reality it would not be a large step.

Authors' response to Broad Comment [8]:

We agree that such an exercise could be useful, but as a science focused and separate topic. However, we wish to limit the scope of this present paper to intercomparisons and issues that arose during this focused objective, in line with the previous comment.

Changes in manuscript re: Broad Comment [8]: None.

Broad Comment [9]: There are many references to the Supplement. It would be helpful if, in each case, the reference pointed to some specific part of the Supplement.

Authors' response to Broad Comment [9] and changes in manuscript re: Broad Comment [8]:
We have checked all references to the Supplement. Most of the references were already to specific sections of the supplement or to the tables. The rest were found to reference uncertainties, in which cases we have added the words "for each specific device". Each section describes the manner in which these were define, and most of these are uniform.

Broad Comment [10]: It is regrettable that the ACP (3.1) and AIDA (3.2) sampling results are in different forms, i.e. per air volume and as site density. Was this unavoidable? Considered unimportant? It comes as a surprise to the reader. Having the two different aerosol processing paths provided some assurance of minimizing problems related to each one. Perhaps there were other operational advantages or limitations as well. Yet, one would have like to see the results in the same format. For inter-comparisons the two data sets provide additional support and that is the main goal of the Workshop. But it isn't clear why the results need to be given in different manner. More careful reading of the paper may reveal the reasons, but perhaps readers (like me) could be helped by a statement of those reasons up front.

Authors' response to Broad Comment [10]: This decision has been discussed in response to previous comments. We understand now that we needed to describe the manner of operations of the workshop that was jointly determined by the workshop leads in consult with participants. The operational protocol determined what we deem as the best means of comparing the different chamber results. It was not meant to leave unrevealed any particular thing about comparisons. Indeed, the decision made to show the AIDA results in terms of active fraction (identical in all respects to a number concentration intercomparison) and n_s was to permit explicit comparison of results from the two sampling periods for the same INP types (even though they were different aerosolization experiments, sometimes on different days). This led to the multi-panel figure comparing active fraction and n_s values from both the AIDA sampling and from the APC, using shading in one panel to underlay one type of data and hopefully not make the figures unreadable. We decided that nothing especially revealing would be lost by not repeating the active fraction plot in each example give in Figures 8 to 11. The timing of sampling largely drove this decision, already mentioned above. There was no way to sample the INP concentration by all samplers, including AIDA, for each aerosol generation experiment.

Changes in manuscript re: Broad Comment [10]: See response to Broad Comment [1]. We definitely want readers to be able to navigate the paper smoothly, so we have additionally added small statements in the Methods section and reminders in the Results section to assist understanding why different formats are used for APC and AIDA sampling results, and how these are brought together ultimately in the later figures.

Broad Comment [11]: The focus on measurements of atmospheric INPs in the recommendations on page 22 is well placed. But ice nucleation measurements are also needed for fundamental understanding of ice nucleation and of the nature of INPs. Another important direction that ice nucleation measurements should explore in future workshops is to perform tests with mono-disperse aerosols.

Authors' response to Broad Comment [11]:
We definitely agree. However, producing monodisperse aerosols for many investigators is yet another logistical challenge. Probably there will need to be a number of different and smaller workshop approaches with different teams.

Changes in manuscript re: Broad Comment [11]: Both recommended topics, and others are listed in an expanded Summary and Conclusions section, appended to the end of the response to Reviewer 2.

Specific items

Specific item 1. page/line 3/32: 'constrain' INP population???
Specific item 1 author response/manuscript change: Removed as superfluous.

Specific item 2. page/line 4/7: The fact that the more general aspects of the project are to be published later is a bit of a problem. A brief description of the three stages would help.
Specific item 2 author response: While we eliminated this material as a possible distraction, we add some words now about the different parts of FIN so that readers can understand the activity and the FIN-02 part of it. The overview article is in preparation.
Specific item 2 manuscript change: In the Introduction, we now add, *"Briefly, and distinct from most previous workshops in its comprehensive scope, FIN sought to perform comprehensive operational comparisons of ice nucleation instruments for sampling calibration type INPs (representative of different atmospheric classes) in a laboratory setting and for sampling ambient atmospheric aerosols in a natural setting. In addition, the component FIN-01 (first study in late 2014) sought to intercompare single particle mass spectrometer instruments that are sometimes used to assess the detailed chemical composition of INPs by sampling the residues of ice crystals nucleated in flow diffusion chambers or aerodynamically segregated from atmospheric clouds. FIN-01 tested these instruments for their determined reference mass spectra on some of the INPs also planned for use in comparing ice nucleation instruments, it compared the different clustering algorithms used by the aerosol mass spectral*

community, and it repeated testing on ice crystal residues. FIN-02, the workshop phase discussed herein, was the laboratory ice nucleation instrument intercomparison. FIN-03, the field phase, was conducted at Storm Peak Laboratory in Steamboat Springs, CO. A final aspect of both FIN-01 and FIN-01 was to provide a minor period within the overall informal gatherings for scientific study that would feature a formal intercomparison of measurements. FIN was a volunteer activity on the part of participants, who agreed to participate to their fullest extent in both the informal and formal components, but were also free to explore new developments. Referees were solicited for organizing and analyzing the results of formal comparisons in FIN-01 and FIN-02. These formal or so-called "blind" experiments were conducted to investigate the degree to which the informal results presented in papers such as this one could be independently reproduced."

Specific item 3. Page/line 4/10-12: The order of these last two sentences of the paragraph should be reversed.
Specific item 3 author response/manuscript change: This section has been rewritten, as noted just above.

Specific item 4. *Page/line 4/22: What do you mean by 'align'?*
Specific item 4 author response/manuscript change: We meant "coordinate." Wording changed.

Specific item 5. Page/line 4/25-32: Goals are well defined. It would be nice if the conclusions responded clearly to each item.
Specific item 5 author response/manuscript change: We now explicitly focus our summary of results around the stated objectives in the new Summary and Conclusions section.

Specific item 6. Page/line 4/33: The intent to define a priori instrument characteristics is somewhat futile. Instruments that participated can be grouped as on-line and off-line but linking that too strongly to dry and wet is too much of a simplification. It would be more effective to describe the methods starting with Fig. 1 and then detail what the instruments do (above water saturation and immersion freezing) and do not do (deposition, condensation-freezing, contact). Indeed, reference to Vali et al, 2015 can help with the definitions.
Specific item 6 author response/manuscript change: The dry and wet designations have been used now in a number of prior published intercomparison studies. Nevertheless, we have refocused all discussion around direct and post-processed sampling systems. The noted sentence now reads, *"These categories are, firstly, instruments operating online or for direct processing of aerosol particles and, secondly, those utilizing collections of particles for subsequent offline or post-processing."* Additional small changes to the entire paragraph remove reference to wet and dry or online and offline, and clarifies what is measured in each case.

Specific item 7. Page/line 5/16: The operating principles are in in Tables 1. and 2 to some extent and are described in 2.1 and 2.2. The Supplement is to describe the detailed implementations of the basic idea in each device.

Specific item 7 author response/manuscript change: Wording is changed. The operating principles are already described in basic publications about most devices, so we attempted to reduce clutter. We write, *"Names, basic descriptions, and general operating principles are provided in Tables 1 and 2, and sections 2.1 and 2.2. Detailed implementations of the basic principles in each device are given in the Supplement sub-sections."*

Specific item 8. Fig. 1: The inclusion of X wt% designation here is unhelpful. The inclusion of this factor complicates interpretation. Indicated range may have been correct for the INPs chosen for test but have no general meaning. There are other materials that cover the whole temperature range with a single wt%. Adding the specific INP for which the graph is valid would add unnecessary detail. I suggest to leave those wt% indications off the graph. Of course, solute effects are ignored here.

Specific item 8 author response/manuscript change: We respectfully disagree, feeling that this point is useful for noting a fundamental difference in how the parameter space is covered by different instruments, even within the same general class. These values were indeed not intended to have a specific meaning except to indicate dilutions of samples to increase dynamic range, a general procedure that is needed for some sample types and droplet volumes. We write in the caption of Fig. 1, *"The blue arrow following the water saturation line in T-RH_w space shows the trajectory of subsequently diluted samples (generically and schematically referred to as X, 0.X an 0.0X weight percent suspensions) of collected aerosols measured by immersion freezing methods. Such dilution is required in many cases for the laboratory samples tested, but the need for dilution or not also depends on the droplet size/volume used."*

Specific item 9. Page/line 6/14: The definition of max RH_w practicable before 'breakthrough' is made difficult by the use of RH_w for many different aspects of the processes.

Specific item 9 author response/manuscript change: We have streamlined this discussion by removing so many references to water relative humidity. We write, *"When the temperature gradient in the growth section is adjusted to generate water supersaturated conditions that activate cloud droplets within the aerosol lamina, the lower relative humidity in the evaporation section shrinks droplets back toward haze particle sizes. This method works up to some high value of RH_w in the growth section whereupon activated cloud droplets survive through to detection, often referred to as the water droplet breakthrough RH_w. The breakthrough value varies with temperature, geometry and flow rate for different devices. Therefore, a single RH_w level in the growth region for breakthrough to occur is not noted in Fig. 1. Instead, results are stated as being associated with specific RH_w values (or % supersaturation values, which equal RH_w-100) that are simply a value that was lower than the droplet breakthrough condition. In some cases, this was the maximum RH_w achievable in the growth region prior to droplet breakthrough."*

Specific item 10. Page/line 6/18-20: Is this caveat not in conflict with the goals defined earlier? May be just say that experiments included in the evaluations all stayed below the breakthrough RH w.

Specific item 10 author response/manuscript change: This is a good suggestion, and one we implement. The sentence is eliminated in preference to the rewrite of the sentences in the comment above.

Specific item 11. Page/line 6/31-33: This sentence should start the paragraph to explain why the choice of 102% needs discussion.

Specific item 11 author response/manuscript change: There was no intention of focusing on 102%, as it was simply put out that as a value that one might not expect to be relevant in the atmosphere. It is relevant for CFDCs. We have removed this sentenced and reorganized/rewritten this paragraph. It was too wordy and somewhat repetitive. This entire paragraph now reads, *"The focus on reporting of flow chamber data at highly supersaturated conditions as best representing proximal immersion freezing behaviors is motivated by recent research and publications. Continuous flow diffusion chamber instruments in general do not presently expose particles to uniform water supersaturations with the precision achieved by cloud condensation nuclei (CCN) instruments. Rather, the transition into the immersion freezing regime above water saturation does not occur sharply in line with the supersaturation calculated for the aerosol central lamina, but rather ensues completely only at higher RH_w as controlled by aerosol particle properties and instrument characteristics. For example, hygroscopicity and kinetic factors control water uptake, chambers have different flow rates and growth section lengths, there is a finite difference in RH_w across the aerosol lamina, and many devices appear (for as yet unclear reasons) often to induce a proportion of all particles to escape the defined aerosol lamina and expose these particles to lower RH_w outside of the intended central lamina (DeMott et al., 2015; Garimella et al., 2017). Hence, higher RH_w is used in these instruments to bypass limitations in achieving CCN activation on the entire particle population, and to increase the condensation rate and thus water content of the formed droplets. The justification is to make the measurement outcome (most particles placed in cloud droplets larger than a few mm prior to freezing) more equivalent to the outcome of cloud parcel simulations in the AIDA cloud chamber (see next section) that are achieved for more typical cloud supersaturations and time scales. This leads to inherent uncertainty in comparing results from these chambers, an issue that we will only acknowledge here, but did not plan as a special focus for this overview paper. In practice, continuous flow instruments processed dry particle samples by slowly "scanning" RH_w from near ice saturation conditions to water supersaturated conditions (see DeMott et al., 2011 for discussion of these methods, and the Supplement Section S.1.2 for a few examples), and these data have been archived from FIN-02. Investigators were then asked to select those data they felt represented the highest (not necessarily maximum) immersion freezing activity it was deemed possible to assess in their RH_w scans, and reported the INP concentrations and RH_w values selected."*

Specific item 12. *Page/line 6/37-7/1: This is problematic. How is this known? The fraction of aerosol activated as a measure of the RH achieved in an instrument? The nominal RH and actual volume-weighted one differ, as just described in earlier part of the paragraph.*

Specific item 12 author response/manuscript change: We believe that the point may have been misunderstood. It was poorly stated. The point is that the instrument lamina RH cannot be interpreted as the exposure RH. As pointed out, this was somewhat repetitive, and so the paragraph has been rewritten. See above.

Specific item 13. Page/line 7/3-7/5: Recommendation here is out of place and heavy-handed.
Specific item 13 author response/manuscript change: We have removed this. It is a clear need for the field, but out of place for this section.

Specific item 14. Page/line 7/7: Shouldn't this 'primary comparison' be defined as a 'goal' up front?
Specific item 14 author response/manuscript change: We believe this was mentioned upfront, and only wish here to point out that additional data are available.

Specific item 15. Page/line 7/11: Quite unclear. Highest vs. maximum. Could the question be said to have been to select the RH point in the scans that the operators thought was most representative for freezing nucleation? Again, this objective should be stated before discussing details of RH variations in the instruments. All those variations are caveats on the validity of the values chosen/used for the comparisons. In any case, isn't it likely that the activated INP were in drops, due to the high hygroscopicity of the aerosol? In that case, the INP's don't experience 102% or whatever supersaturation. Clearly, these considerations need to be addressed here.
Specific item 15 author response/manuscript change: All of the INPs are not in the drops in CFDCs up to some RH that exceeds the expectation, and that is the point of the discussion, that every group operating such instruments has noted this. The suggestion is well taken though, and this section has been rewritten, as noted above (specific item 11 response).

Specific item 16. Page/line 8/5: Is this because only ice crystals forming from frozen droplets on the surface are counted? How certain is that? Same issue as in previous point.
Specific item 16 author response/manuscript change: We are not sure that we fully understood this comment, but hopefully some added words and reorganized discussion help here. We were only remarking that data will be included from the substrate-based diffusion chambers up to reported values of 101-102% RH with respect to water, but acknowledging that the equivalence of INP response to immersion freezing remains under evaluation, especially as this may depend on particle loading. We did not wish to gloss over this issue, only to limit extensive discussion of it when we are referring to two particular instruments. We have added overall to this discussion, including some past references, at least ones that reference the substantial literature on this topic, as note by the reviewer. See next comment.

Specific item 17. Page/line 8/7: " ... limits reduction of ..." is unclear. The influence of vapor competition, and its dependence on INP density on the substrate has extensive literature, and perhaps should not be glossed over so easily. The operational definition used here for making the comparisons deserves more explanation. The concluding sentence on 8/13-15 is unclear - do you mean cases when vapor competition was only inferred to have been present but not evaluated?

Specific item 17 author response/manuscript change: Upon review, only one of the submitted cases was removed from consideration by the DFPC group, and this related to potential heating (of the paraffin) impacts on Snomax activation, still under investigation. Hence, we have revised this section to emphasize matter-of-fact discussion of the two devices, and to mention that we are including them in this intercomparison of immersion freezing despite a continuing need to evaluate their ability to represent immersion freezing. We write, *"These two methods were developed to measure condensation freezing and deposition ice nucleation modes from below to slightly above ~~at~~ water saturation and below. The thermodynamic path of measurements using these instruments is the same as for the continuous flow diffusion chambers in Fig. 1 (red lines), but typically terminate 1-2% RH_w above water saturation. Both devices were designed with the intention to overcome the so-called "volume effect" on freezing (e.g., Bigg, 1990; Schrod et al., 2016 and references therein) which describes the underestimation of INPs that can occur when processing particles on a substrate in a diffusion chamber due to vapor pressure reduction by the first particles freezing, especially when larger volumes are collected that result in larger numbers of INPs per surface area of the substrate. The FRIDGE instrument seeks to limit this effect using a low-pressure diffusion chamber to enhance vapor deposition over particles collected onto silicon wafer substrates, while the DFPC instrument follows the methods of Langer and Rodgers (1975) to focus a flow of humid air over filter substrates, and using the best practices outlined by Bigg (1990). For both devices, attempts were made to limit particle collections to shorter times during FIN-02, in order to keep particle loading light on the substrates. An additional fundamental differences between FRIDGE and the DFPC is the use of the filter substrate in the latter case, which is placed on a paraffin layer that is heated to establish thermal contact with a cold plate prior to ice nucleation measurements. The uncertain difference between condensation freezing and immersion freezing mechanisms (Vali et al., 2015) argues for an evaluation of both methods by including their results obtained at near water saturation in this intercomparison as representative of proximal immersion freezing."*

Specific item 18. Page/line 8/8: '...were allowed to ..' inserts a hierarchical tone that is unnecessary. An objective tone would be more appropriate. The working arrangements of the workshop are not of interest to the reader. Else, make it clear and explain in the Introduction that there was some sort of checks and balances arrangement in order to improve the final result. I don't know if there was or not. See also comment on 7/3-5.
Specific item 18 author response/manuscript change: We have attempted to remove any hierarchical tone, and revise as noted above. Again, this was the informal component of the workshop, and the intent was to offer the best chances to compare instruments on a fair basis. We included as much of the data as was possible, and truly tried to limit and omissions.

Specific item 19. Page/line 8/14: "wet suspension" -- ?? wouldn't liquid suspension or water suspension be more direct? Actually, the first two words could well be omitted and the sentence say "Measurements of immersion freezing ...."
Specific item 19 author response/manuscript change: The latter suggestion would ignore that immersion freezing describes a process and not a stock measurement method, in our opinion. We here use *"immersion freezing*

measurements of collected particles suspended in water." In other cases, we take the tip to simplify to *"water suspension*".

Specific item 20. Page/line 8/31: " ... wet suspension groups ...? -- one can deduce what is meant but it could be said better
Specific item 20 author response/manuscript change: We have revised simply *to "Most groups using liquid suspension freezing methods group shared common samples…"*

Specific item 21. Page/line 9/1: "... basic methods ... " may need to be re-considered if changes are made in response to the first item in this review.
Specific item 21 author response/manuscript change: Done, as mentioned above.

Specific item 22. *Page/line 9/29: suggest leaving out the word 'mode'*
Specific item 22 author response/manuscript change: Done.

Specific item 23. Page/line 10/3: suggest replacing 'wet dispersion of particles' by something like "aerosol generation from aqueous suspensions" or "particle dispersion from aqueous suspensions" or "dispersion of particles via spraying and drying"
Specific item 23 author response/manuscript change: We use, *"Aerosol generation from aqueous suspensions..."*

Specific item 24. Page/line 10/32: Was the depletion of aerosol in the APC by sample withdrawal avoided or neglected? Also, how justified was it to assume the decrease in concentration to be valid for all sizes?
Specific item 24 author response/manuscript change: Depletion of aerosol by sample withdrawal was in fact the primary mechanism for reduction in concentrations. The fits encapsulate this primary mechanism and wall losses. Brownian diffusional losses to the walls was of course limited due to particle mode size being typically a few to several tenths of a micron in diameter. Hence, particle sizes were in the range where large differences in reductions as a function of size over time were not expected. This was a reason for showing a figure to include two different size ranges. We found that using total particle numbers versus using the concentration of particles > 0.5 μm led to no more than a 30% difference in the fractional losses of particles over time during the initial fills with higher total aerosols. This is evident in Fig. 3. During the time period of direct particle sampling, the particle losses losing the same size ranges differed by no more than 10%, also evident in the figure. Hence, we assume a similar range of uncertainties on correction factors to bring the direct sampler INP concentrations into line with the period of sampling by post-processing devices. We have added statements in these regards. *"These loss processes were dominated by the drawing of air from the APC by samplers, replenished in all cases by clean synthetic air."* Also, *"While we focus in the following discussion on quantifying the decay of total (CPC) particle numbers in order to correct INP number concentration data during the direct sampling periods ("online" used as shorthand in equations) for equivalency with the prior post-processing collection periods ("offline" used as shorthand in*

equations), we noted (not shown) differences ranging from only 10-30% in fractional loss rates when instead using particle numbers in the larger size range (>500 nm) to characterize particle number decay over time in the APC. These relatively minor differences, evident in Fig. 3, are consistent with the limited physical loss mechanisms existing for particles with mode sizes as shown in Fig. 2, and with limited numbers of supermicron particles that might be subject to sedimentation."

Specific item 25. *Page/line 12/30: perhaps the intention was to put the words 'basis for" in the sentence, to read : .... primary basis for the comparison of ...*
Specific item 25 author response/manuscript change: Done.

Specific item 26. Page/line 12/31: replace 'most' by 'largest number' ??
Specific item 26 author response/manuscript change: Done

Specific item 27. Page/line 13/18-20: The sentence is somewhat garbled.
Specific item 27 author response/manuscript change: Given the extended earlier discussion, we have removed some discussion here and have revised this sentence to read, *"We note again here that some scatter may occur in the online methods due to investigators deciding in each case at what RH_w above 100% to report data for representing immersion freezing."* We may note to the reviewer that this topic will be featured within the referee paper on the formal workshop component of FIN-02.

Specific item 28. Page/line 14/3-22: These comments would be better placed in a Discussion, not in the Results section.
Specific item 28 author response/manuscript change: We have renamed the section to account for the inclusion of discussion.

Specific item 29. Page/line 15/12-14: I think it is too uncertain to explain results in terms of the numbers of proteins. Snomax contains full cells and evidence for separation of the INP material (protein) from the cell wall is unclear. Leveling off the INP curves shows a limit in the number of INPs as a fraction of the total number of cells. This is a matter of expression of the INP protein under induced conditions and of the processing of the bacterial culture. Again, the emphasis here should be on the agreement among measurement systems. Interpretation of the shape of the INP curves is a matter for Discussion (see Broad comments 7 and 8).
Specific item 29 author response/manuscript change: We were attempting to explain the behaviors in terms of discussion present already in past literature, but understand the point here and agree that we should omit this statement as tangential to the overall comparison.

Specific item 30. Page/line 15/25: Reference here to first-freezers etc. is confusing without more detailed knowledge of what is being discussed. This type of error analysis for specific devices should be part of the apparatus description (cf. Broad comment 6).
Specific item 30 author response/manuscript change: We were not discussing an error analysis in this section, but pointing to the fact that in some experiments, at least two of the freezing methods detected discrepancies in freezing conditions of their samples. We were also pointing out that such a result is consistent with the impact of handling on the first freezers detected on the basis of other published work.

Specific item 31. Page/line 15/32: Supporting Information =? Supplement. Where is the material referred to in this line?
Specific item 31 author response/manuscript change: Section S.2.2, stated now, to point specifically to the CMU methods.

Specific item 32. Page/line 16/10: Why wasn't the fraction expressed as active INP versus total number of potential INP particles introduced? Maybe that is what is meant, but I am not sure from the wording given here.
Specific item 32 author response/manuscript change: That is what was intended. The fraction was INP over potential INP for each type, representing a different experiment in each case. Reworded for clarity as, *"These data were first analyzed as active fractions, which is the number fraction of all particles freezing when normalized to the total number concentrations of particles (potential INP) present at the onset of expansions."*

Specific item 33. *Page/line 16/16-17: The conversion to active site density is the same for all measurements, provided particle sizes are known. Why is this introduced here? (cf. Broad comment 10)*
Specific item 33 author response/manuscript change: The active site density approach was introduced in Methods. It is mentioned here because we are about to present figures including it. The reasoning for this was provided in response to Broad comment 10. It was not possible to compare AIDA and APC experiments in any manner except on the basis of active fraction or n_s. We choose to show one active fraction plot and then proceed to reduce figures by showing only n_s. By overlaying the AIDA experiment ns data on n_s calculated from the APC experiments, we seek to tie the two series together.

Specific item 34. Page/line 20/28-31: Do these two sentences refer to the same or two separate findings?
Specific item 34 author response/manuscript change: Rewritten for clarity. There are two points. The steep activation behavior with temperature exacerbates discrepancies. It was also the steep slope of the FS02 activation versus temperature that led to the finding that CFDC cooling in the evaporation region can express "late" activation of ice crystals that remain at small sizes and should not be attributed to the set point temperature of the lamina in the instrument. We write, *"The steep activation behavior of the FS02 also led to the finding that when sampling such INPs, cooling to achieve evaporation in the exit section of a CFDC can express "late" activation of ice crystals that*

remain at small sizes and should not be attributed to the set point temperature of the instrument growth region. This may be an issue primarily for laboratory measurements of such INPs, since most natural INP T-spectral slopes…"

Specific item 35. Page/line 21/21: Some numerical values for the errors discussed would be useful.
Specific item 35 author response/manuscript change: We are not sure how to answer this request, since it deals with a topic that was not investigated. Namely, we did not purposely attempt to create size distributions of INPs that would challenge instruments with size cuts. Furthermore, to suggest a range of errors that would occur in the atmospheric scenario, we would need to know a typical size distribution of atmospheric INPs. This could vary tremendously. The topic is appropriate for an ambient measurement intercomparison. It could be explored in a laboratory setting, but as mentioned, we did not explore it during FIN-02. Hence, we make no changes here.

Specific item 36. Page/line 21/23: 'de-agglomeration' has a simpler alternative: 'breakup' and removes the implication that all large INPs are aggregates of many smaller ones
Specific item 36 author response/manuscript change: Very good point, and accepted.

Specific item 37. Page/line 21/28: " ... full immersion of all particles in the same liquid volume .." is too vague to focus on, as other factors like time in suspension may also come into play. Also, are the differences beyond the error bars of the PIMCA-PINC and cold-plate methods?
Specific item 37 author response/manuscript change: We have added on breakup, sedimentation and active site alteration in bulk suspension as possible impacts on freezing spectra compared to PIMCA-PINC single droplet results. As for the last question, error bars are shown on all plots, so the answer is yes at the level of the experiments performed. If referring to the overall uncertainty evident in experimental results, it is harder to say. This topic would benefit from more overlap in the measurement regimes of the two methods for more instruments, but we have noted that PIMCA-PINC is restricted to assessing higher fractions freezing. We write,*"Reasons why the offline bulk immersion freezing methods do not always agree with PIMCA-PINC may relate in some unresolved manners to factors at play during extended bulk immersion, such as breakup, sedimentation, and alteration of active sites. "*

Specific item 38. *Page/line 21/32: why is that need artificial?*
Agreed. It is artefactual, as in, not fundamental to the original operating principle.
Specific item 38 author response/manuscript change: We drop the artefactual part of that statement.

Specific item 39. Page/line 22/4: '…uniformly capture activation ...' is awkward wording
Specific item 39 author response/manuscript change: We have changed the word to *"equivalently"*.