Response to the reviewer 1 of the manuscript: "Comparison of different droplet measurement techniques in the Braunschweig Icing Wind Tunnel" by Inken Knop et al.

We thank the reviewer for the encouraging comments. We highly appreciate the outstanding knowledge and experience of the reviewer. Our responses to the suggested revisions can be found below. Changes in the manuscript are marked in red color.

This manuscript describes an icing tunnel test facility that may be of interest to the atmospheric science community. The fundamental operation of this system is similar to existing icing tunnels, however, there is no comparison to icing tunnels that have been in existence for nearly 50 years. Of particular interest is the icing tunnel in Ottawa, Canada that is operated by the National Research Council. A description of the NRC icing tunnel can be found at https://nrc.canada.ca/en/research-development/nrc-facilities/altitude-icing-wind-tunnel- research-facility, and early papers describing experiments in the tunnel are by Strapp and Schemenauer (1982) and King et al. (1985). The NRC icing tunnel has unique advantages over the Braunschweig tunnel, specifically the NRC tunnel is capable of particle speeds up to 100 m s^{-1} and altitudes to 40,000 ft (12 km). The Braunschweig maximum particle speed of the Braunschweig tunnel is 40 m s^{-1} and it has not capacity to simulate altitude. This is unfortunate since research aircraft fly at various altitudes and all large aircraft fly at speeds that are at least twice the maximum speed that the Braunschweig tunnel can produce. The manuscript needs to discuss how the limitations of the Braunschweig tunnel influence their results.
Response: There are indeed many icing wind tunnels worldwide. The authors are well aware of the NRC facilities and have many excellent collaborations with the research teams in Canada. However, the goal of our paper is not to perform a tunnel-intercomparison exercise. Instead, we focus on the intercomparison of several droplet sizing techniques in one tunnel, the Braunschweig Icing Wind Tunnel.
The Braunschweig Icing Wind Tunnel has many applications, not only limited to aircraft icing. Also the presented measurement techniques are not only limited to applications in civil aviation. Therefore, the maximum speed of the tunnel and its sea level pressurization do not impede the validity of the results presented.
Nevertheless, we want to mention that numerous scaling methods based on similitude of geometry, droplet trajectories and the impingement heat transfer have been developed to scale the model and test conditions to simulate the conditions beyond the IWT abilities. A comprehensive description of the scaling methods can be found in Anderson (2004). Scaling one of the parameters requires changes in other parameters to maintain the similitude, often it is not possible to concurrently match all the similarity parameters, especially the pressure which in most of the tunnels cannot be controlled. The AIWT operated by NRC and CIRA icing wind tunnel have the ability to control the pressure to simulate the high altitude icing conditions. Anderson (2004) reports the factors that highly are highly sensitive to pressure have only a limited influence on the parameters that influence the icing the most. From the icing data available at AEDC Barlett (1988) states the influence of pressure on icing is insignificant, the joint NRC CIRA experiments to study the effect of pressure (Oleskiw et al. 1996) showed "relatively small changes in the forward-facing portion of the profiles".
We have modified the paper to include some of the above mentioned comments, and also mentioned the NRC facility in the introduction section.

Understanding discrepancies between drop concentrations and drop size distributions (PSDs) measured by various probes is of critical importance for cloud physics, albeit icing studies rely more on bulk quantities such as MVD and LWC. MVD and LWC are presented and discussed in great detail, but there is almost no quantitative discussion of drop concentrations and PSDs from the PDI, FCDP and Shadowgraphy instrumentation. The manuscript needs to include additional Figures that show

correlations between reported drop concentrations and PSDs measured by PDI, FCDP and Shadowgraphy.
Response: The response of these techniques for a finer spray (MVD 14.5 µm) and a coarser (MVD 33.8 µm) is understood from the bin-wise droplet counts and the corresponding cumulative mass fractions in the additional plot in Figure 9. The trend of the PSD for all the three methods is almost similar up to 50 µm, the FCDP measured count is in general almost an order higher than the PDI. Although only a few droplets above 30 µm are found with shadowgraphy, their weight is enough to deviate the cumulative mass curve from the others.
The droplet concentrations for FDCP and PDI are plotted in Figure 12. For the shadowgraphy technique, the droplet density is not obtained because of the difficulty in defining the probe volume. It can be seen that the FCDP gives slightly higher number concentrations The new plots are discussed in detail the manuscript.

The PDI measures LWC using eqn. 4, which proportional to the product of total drop concentration and corrected volume mean diameter. The manuscript should also show how LWC compares using this technique with LWC computed by integrating the complete PSD. The comparison should be shown as a function of MVD and drop concentration.
Response: This could be an additional interesting consideration. However, it would only broaden the knowledge about PDI. In contrast, the focus of the paper is on the comparison of measurement techniques rather than the detailed investigation of a single measurement technique. Therefore, we refrain from integrating the proposed tests for LWC calculation of the PDI into our investigation.
Furthermore, the LWC computed using equation 4 has shown a good agreement with the results of the WFR, so the manufacturer default (equation 4) seems suitable for our application.

The manuscript compares LWC measurements from the FCDP, PDI and RCT using the WFR as a standard. It is implied that LWC using WFR as a standard is very repeatable, on the order of 7%. In Section 4 a statement refers to Section 3 as justification for this, but far as I can tell in Section 3, this repeatability comes from the literature, not from actual tunnel tests. Yet, I assume there were LWC repeatability tests, similar to the MVD tests shown in Fig. 5, so please point out where I missed the LWC repeatability tests or include Figures showing results from them.
For all the experiments involving the LWC computation, the water flow rate is recorded and the corresponding LWC is computed. In total LWC from WFR is computed for more than 400 individual cases, of which several sets have been repeated to determine the variance of the water flow meters. The mean coefficient of variation of these repetitions is calculated to be 7% as reported in section 2.3.

Now, that said, based on the large amount of scatter shown in Figs. 12 – 14, either the tunnel flow characteristics or the measuring techniques, or both, appear to be contributing much more variability to LWC than the 20% Figure quoted in the text. It would be useful to show a complete uncertainty analysis for the WFR and test instruments measurements, but this is likely to be outside the scope of this paper. However, the manuscript should address tunnel and instrument LWC uncertainties in a more rigorous manner, not just quote the literature.
Indeed, a full uncertainty analysis based on the detection physics of each measurement technique is out of the scope of our manuscript. Nevertheless, we want to provide some estimates here. The uncertainties of the tunnel flow characteristics and the measurement techniques are both acting on the results presented in Figs. 13-15.

Let us therefore comment on the tunnel characteristics first. Here, the aerodynamic performance of the tunnel and the liquid atomizers that produce the droplet cloud need to be considered.

1) Aerodynamics: The repeatability of the wind tunnel and nozzle input conditions are studied and plotted in Figure 4 The precision limits for these variables for a sample run are also reported in section 2.3. The aero-thermal characteristics of the tunnel have already been calibrated as per the guidelines of SAE ARP 5905 with the recommended instruments and uncertainties which is now included in the manuscript. Thus, the temporal stability of the tunnel is guaranteed.

2) Liquid Atomizers: the repeatability of the spray can be appreciated from the plots in Figure 5. The temporal stability can be seen in Figure 6.

The atomization physics is highly dependent primarily on the supply pressures of water and air and the operating duty cycle. The fluctuations of these critical parameters lead to a higher uncertainty in the PSD and the LWC. To better estimate the uncertainty of the spray, additional data from another new spray system (not part of this manuscript) shall be mentioned here. The new system is equipped with a high accuracy Coriolis flow meter (accuracy 0.2%), the data was used to formulate an empirical form for the LWC (the variable being the input conditions to the nozzles), the 95% confidence interval of the model with the measurements is considered as the systematic bias of the model, the highest fluctuations of the pressure are considered as precision terms and the root-summed-squared (RSS) uncertainty computed over a wide range of operating conditions was found to be 0.045 g m^{-3}, yielding an total uncertainty of the spray LWC of 10%. This value is slightly higher than the repeatability characteristics, which are given by the coefficient of variation of 7% of the thermal mass flow meters used in the present study.
Given this tunnel operational constraint of creating an LWC with uncertainty of 10%, the fluctuations in Figs 13-15 beyond that value can be attributed to the uncertainties of the individual measurement techniques.
We mention these uncertainty considerations in Section 2.3 in the revised manuscript.

The sample conditions of these tests (droplet concentrations sometimes exceeding 2000 cm^{-3}) are typically only found in polluted environments. This, plus the slow droplet speed, limit the usefulness of the results of these experiments. These limitations need to be discussed in detail in the manuscript.
Response: It is true that the conditions used in this experiment only covers the lower boundary for what e.g. the FCDP is intended for, particle speed wise and at the same time uses large droplet number concentrations, which increases the likelihood of coincidence.
The wind tunnel is also for other applications than simulating flight conditions e.g. icing on wind turbine blades etc.
Nevertheless, the overall special conditions and limitations of the probes are going to be discussed in more detail in the new draft.

The poor sampling statistics for drop diameters > 30 microns (Figs 4 and 7) definitely introduce uncertainties in the LWC results that need to be addressed more rigorously.
Finally, it should be pointed out that without a rigorous uncertainty analysis of the absolute accuracy of the tunnel, all of the quoted accuracies are not absolute, but instead relative. That is, in addition to the random error associated with tunnel properties, there is some degree of undetermined bias error that is not considered. This needs to be emphasized in the manuscript, albeit, hints of this are included in some of the references cited.
Response: The higher bounds of the PSD for some conditions in BIWT can be approximated for example with a Langmuir D type distribution where D_{max} is 2.2·MVD. The expected D_{max} for the spray in Figure 3 (prev Fig .4) is 26.2 µm. The measurement shows droplets above 28 µm contribute less than 0.05% of the total volume agreeing with the expectated thus confirming the validty of the measurement. When the droplet count above 28 µm is doubled the change is in D_{30} is neglible from 9.38 µm to 9.48 µm that also results in negligible change in LWC. For this sample, the vicinity of 5.5 µm dominates the mass curve.

However, the presence of larger droplets in a small sample (1000 droplets) has a perceivable change in the MVD thus the D_{30} (Figure 6 right). As the sample size is increased (above 10000), the presence of the large droplets is reliably accounted and the uncertainity in MVD and D_{30} and LWC will be reduced as shown in Figure 6 left. Accordingly, all the PDI measurements are made with at least 20000 droplets per sample (for individual cases with low data rates at very low LWC) and in average approx.. 60000 droplets. With the FCDP the samples consist of at least 35000 droplets and in average of approx. 60000 droplets.

On the contrary, if the primary mode is not adequately sampled like the shadowgraphy data in Figure 9 left, any change in the count on the tail end of the distribution would alter the D30 and thus the LWC significantly. However, in the present paper, no LWC estimates are made from shadowgraphy.

The inherent complexity makes it impossible to derive a theoretical model for the PSD and lack of any other means in the present project to determine the actual PSD makes it difficult to quantify the bias. Albeit, the repeatability is higher as shown in Figure 5. Therefore, all of the measurements have some amount of unaccounted bias, but with a high precision thus a quality inter-comparison of the methods can be reliably made. These aspects are discussed explicitly in section 2.

Specific Comments:

1. Introduction

Page 2: When mentioning the cloud probes used by Ide (1999) and Cober et al. (2001), the manuscript should describe the resolution and size range of these probes so results can be compared with tunnel results.

Response: Ide (1999) performed the icing experiments in NASA IWT with MVD in the range 10 to 270 µm and velocities 22 to 112 m/s. LWC calculated by integrating the PSD spectra obtained from a combination of FSSP and OAP was reported to be significantly higher (1.2 to 2.7 times). than the LWC measured with icing blade an RCT. The large deviation was attribted to spectral broadening and coincidence errors. This is discussed in the manuscript in the introduction and in section 4.2.

The measurement ranges of the probes used by Cober et al. (2012) in the flight test campaign is also included in the manuscript.

2. Experimental Setup

Add a table (perhaps as a supplement) indicating the mean operating conditions for all of the data sets presented in this manuscript (Velocity, Temperature, RH, Air Pressure, Water Pressure, Water mass flow). This would be helpful to understand the scope of conditions for each type of drop measurement system. Also, list the number of data sets collected for each drop measurement technique (e.g., FCDP: 100 samples at 20 m s^{-1}, 200 samples at 30 m s^{-1}, 300 samples at 40 m s^{-1}).

Response: Certainly, this table would benefit in establishing the validity regions of these measurement and the degree of statistical reliability. The test conditions of all the runs are uploaded as a supplement and a short summary table is included in the paper.

Based on the data plotted in Fig. 11, it appears that the PDI datasets greatly exceed those of any other probe. This could be due to the high particle rejection rate and very low sample volume of the PDI. Please explain.

Response: LWC from WFR is computed for all the measurements along with PDI, FDCP and RCT and has therefore the largest number of measurements.

In this project PDI is used for reference calibration of the tunnel therefore more measurements are made with PDI. The reliability over a wide range of droplet sizes and the low acqisition time needed for a reliable sample enabled more than 300 measurements including repetitions.

On the other hand, due to the limited detectable droplet size range of FCDP and a limited time availability of the probe, only 34 valid LWC measurement were made with FCDP. Also, the measurements with RCT were limited to 37 different spray conditions due to the long measurement duration and the necessity to maintain extremely cold temperature (<-18°C).
The large disparity in number of measurements is purely from the above and not related to the low sample volume.

3.1. PDI
In this section it is noted that "The PVC has the greatest effect on the smallest size classes. Their influence on the LWC, on the other hand, is very small." We have performed an independent analysis of the FCDP data set from this icing tunnel and determined that the peak of the particle mass size distribution is between 8µm and 10µm. Therefore in this study the vast majority of drops are very small (see PDI PSD in Figure 4 for confirmation where the mode is ~6 um). Given these concerns, the manuscript should include more details of the PDI small drop corrections and better quantify the errors. From Chuang et al 2008: "At very small droplet sizes, diffraction can become significant relative to refraction, and lead to oscillations in the φ versus d relationship at the smallest drop sizes, primarily in the size range below 4µm, but with some effects up to ~8µm."
Response: Thanks for invigorating our discussion on this aspect. This will be critical for extremely fine sprays where the mode is observed below 8 µm. The droplet size in PDI is obtained from the linear relations between the phase shift and size derived for a predominant reflection or refraction mode based on geometrical optics (Ofner 2001). Below 5 µm, the validity of the geometric optics tends to cease and the diffraction becomes significant leading to erroneous measurements if the linear relationships are used as mentioned in Chuang (2008). Bachalo and Sankar (1996) reported the uncertainty resulting from these oscillations to be under ±0.5 µm.
Chuang et al. propose using a large off axis angle for attaining higher accuracy of these small droplets but at the expense of the limiting the upper size. A discussion of the above is now included in the manuscript in chapter 4.1.

3.2. FCDP
The description of the FCDP sample volume is fairly convoluted. It is sufficiently described as SV=SA*TAS, where the SA is defined by calibration for a fixed qualification criteria. The SA is defined by laboratory calibration like that described in Faber et al. (2018). See the Table's section of the review for more details.
Response: We agree that the description of the sample area can be facilitated.

In paragraph 3 it should be noted that the CDP and FCDP have similar operating principals, but the improved optics and electronics in FCDP allow for accurate sampling in higher particle concentrations (see comments on Table 2 below). The FCDP also differs from the older FSSP-100 probe in that the qualifier detector uses a slit aperture (200µm x 800µm), which was first introduced on the FSSP-300 probe with data described by Brenguier et al. (1998).
Response: We see that the FCDP is equipped with state of the art electronics and an advanced optics, compared to the CDP. When we reference system and operating specifics of the CDP, we do this as to show capabilities of an example for forward scattering spectrometer probes, without the intention to attribute its specifics to the FCDP. It is rather to put this measuring technique in general in comparison to the other techniques.
Furthermore comparable specific references solely applicable to the FCDP have not been found by the authors.
In the revised manuscript the difference between both systems, CDP and FCDP will be emphasized.

Lance et al. (2010) note that accurate sizing of the CDP instrument to ~200 cm^{-3} before being influenced by coincidence. However, improvements in the CDP (new limiting apertures) increased accuracy such that only 27% undercounting is estimated at concentrations of 500 cm-3 (Lance 2012). This level of uncertainty is still problematic. The FCDP was designed to incorporate the improvements of the CDP as well as reduce particle coincidence (the dominant source of error) by reducing the laser beam waist from 200µm to 80µm). Flight tests indicate reasonable agreement for LWC between the FCDP and hotwire probes for small droplet concentrations as high as ~1000 cm-3. The conditions in this study exceed these typical atmospheric conditions, so the FCDP uncertainty for these high drop concentrations range is not well described.

Response: In our citations we address mainly sources related to CDPs. So that repeated references to the (older) CDP are made throughout this chapter. A thorough comparison of both probes specifics and a comparison to other forward scatter cloud probes lies beyond the scope of this paper. But we agree that the major improvements of the FCDP versus the CDP, namely the electronics and optics should be addressed on the course of this chapter, especially when it comes to coincidence.

A usable citation, where confidence towards droplet concentration measurements ~1000cm-3 is expressed would be of help. The regime in which the FCDP is operated in this experiment will be discussed.

In paragraph 3, the authors cite up to a 50% uncertainty from Baumgardner 2017, but it should be noted that this quoted uncertainty (10 to 50% for light scattering probes), includes "Mie ambiguity, collection angles, coincidence, nonsphericity and shattering." In this study all droplets are assumed to be spherical, and shattering is minimal for the FCDP, so the 50% uncertainty does not apply here.

Response: It would be better to emphasize that the quoted citation from Baumgardner, 2017 is a maximum value for generic particle forward scattering probes, including all limitations for this measuring principle, including internal and external factors, which contribute to the cited up to 50% uncertainty. It is a good hint from the reviewer to elaborate the composition of the overall uncertainty and to point out what really applies for this probe and this experimental setup.

3.3. Shadowgraphy

Overall the Shadowgraphy technique is poorly described. More details of the instrument and post-processing are required such that the test could be replicated and verified by another group.

Response: The description is improved in chapter 3.3 with the details of the optics and light source. A description of the calibration is made. The DoF and border correction terms are discussed. Some recommendations are made from the experience. A description of the post processing is also made in the manuscript. The equipment specifications are also appended in the corresponding table 3.

In the last sentence from this section the data inter comparison is considered" almost identical." This statement requires quantification.

Response: In total 35 measurements are made with shadowgraphy, only 20 have individual conditions. The remaining15 are either repetitions or measurements with change in velocity for the same spray settings. Sixteen samples have MVD_{PDI} below 35 µm, which show the correlation of $MVD_{Shadow}=0.96 \cdot MVD_{PDI}$. The details on the measurement points are included in chapter 3.3 and the test matrix is added as a supplement.

3.4. Rotating Cylinder Technique

Stallabrass (1987) should be Stallabrass (1978)

Corrected

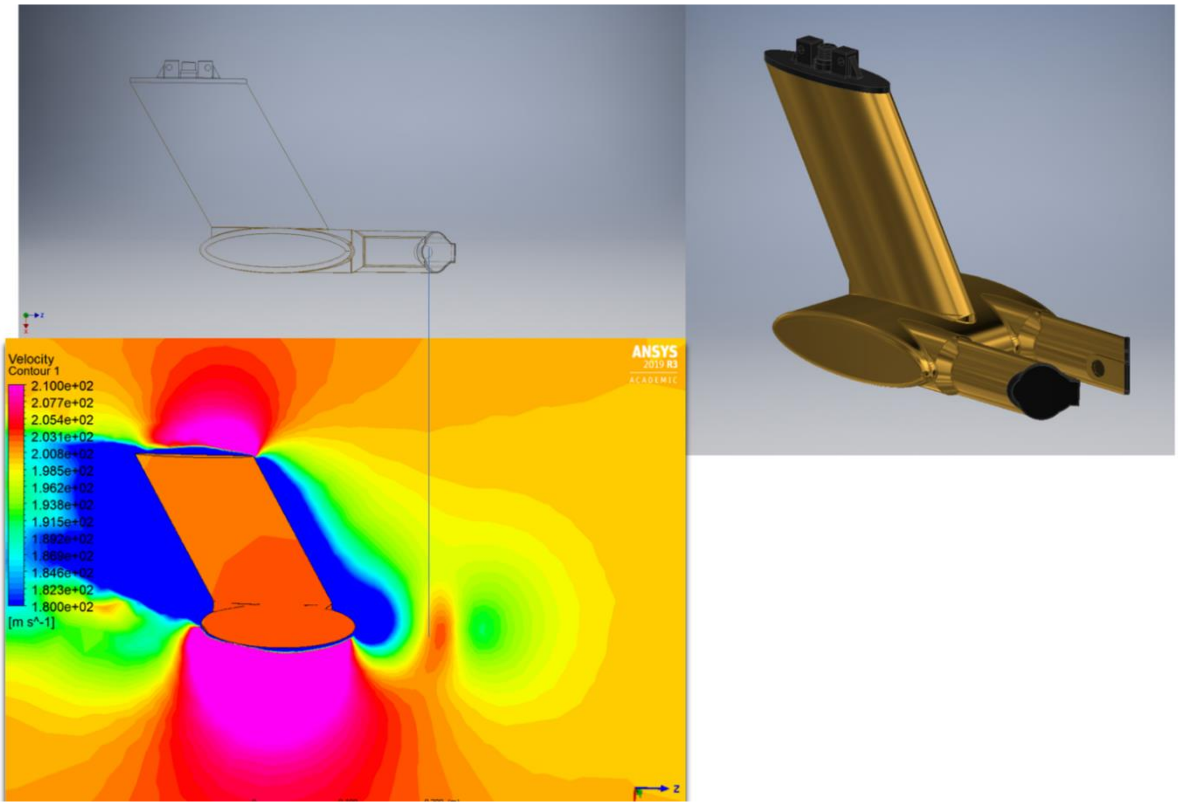
4.1. Repeatability

See General Comments.
The discussions of the wind tunnel temporal stability and its repeatability can be found in the new chapter 2.3. The analyses of the overall combined wind tunnel and measurement setup precision is now added at the end of each section in chapter 3.

Paragraph 2: “see section 4.1” within section 4.1, should be “see section 3.1”.
Corrected
Paragraph 4: “see section 4.2” within section 4.1, should be “see section 3.2”.
Corrected

4.2. Comparison
Paragraph 3: “A low sensitivity of the FCDP to larger particle sizes (> 30µm) may cause or contribute to the measured deviation of the FCDP with respect to the PDI for large droplets.” Is there evidence that the FCDP has a low sensitivity to larger drops? If so, provide a reference. The number of sampled drops is relatively low, due to the small sample volume and low concentration of larger drops, but this is true for all single-particle devices, including the PVI. Also, with long runs in an icing tunnel this should not be an issue, assuming the tunnel properties are repeatable, as claimed in Section 4.1.
Paragraph 3: “The transit time filter applied to the FCDP data during post-processing to reduce coincidence causes a rejection of droplets that have a too long transit time compared to the mean reverence [sic] velocity and thus reduces the droplet size spectrum evaluated as valid by large droplets.” Is there any evidence to support this assertion? If so, please provide the evidence, a reference, or sound physical explanation.
Response: In this case the authors refer to an effect not to be attributed to the FCDP probe itself, but to a slip of larger droplets within the airflow, compared to smaller droplets and the subsequent application of the transit time coincidence filter. Since droplet speed of larger droplets seem to be slightly lower than the airstream and in addition their sample number is overall fairly low, we have observed the tendency that the current transit time filtering for coincidence, based on the given TAS assumed as the particle airspeed, can lead to discarding of genuine counts of larger droplets, which deviate more than 25% from the C1C3 distribution, based on the given TAS.

Paragraph 5: “According to Lance et al. (2010), an additional source of error of the CDP might be the external geometry of the probe, which can alter the measured cloud particle size distribution.”
This statement does not apply to the FCDP because it has “anti-shattering tips” that minimize droplet splashing, whereas the CDP Lance used did not (at that time) have anti-shattering tips. The CDP can be equipped with anti-shattering tips now.
Response: We have to thank the reviewer for pointing out the fact that both CDP (in its initial design) and FCDP differ among other points in probe geometry. One major advantage of the FCDP’s shape is the application of anti-shattering tips. The challenge of droplet splashing is thus reduced. Nevertheless exposing in-situ probes into a droplet laden airstream alters the flow locally. As can be seen from the CFD-Simulation below and Spanu, et al. (2020), Weigl et al. (2016). When comparing different measuring techniques, this is an important factor to be mentioned, when discussing measurements.



4.3. Comparison of LWC measurements

See General Comments, also:

Paragraph 6: “This can only be explained by higher particle number concentrations measured by the FCDP compared to the PDI.”

Please provide particle concentrations and size distributions for the FCDP and PDI for the relevant wind tunnel datasets.

Response: The droplet number densities acquired by FCDP and PDI show a good agreement, that is shown in figure 12. In average the FCDP gives little higher concentrations than the PDI, what might be a possible explanation for the higher LWC results of the FCDP. The compared size distributions are shown in Figure 9. Both new figures are discussed in Chapter 4.1.

Following the example in Figure 7, the authors should indicate when the PDI and FCDP sampling statistics are poor (<100 particles per bin) and possibly remove these data from consideration.

Response: Our typical wind tunnel droplet size distribution has in almost all cases a long end with only very few large diameter droplets, as discussed in section 2.4. This can also be found in the literature (Rudoff et al., 1993; McDonell and Samuelsen, 1996). The number of particles per bin is as well a question of the bin size. In figure 7 we choose a bin width of 2µm. Increasing the bin width would automatically lead to higher droplet counts per bin. With a minimum droplet count of 10000 per measurement we present in the paper only measurement data with statistically secured MVD values.

5.0. Summary

Page 17 Paragraph 2: “The characterization of cloud droplet distributions with particle sizes > 100µm poses new challenges for droplet measurement techniques.” Some Optical Array Probes are well suited to particle measurement in this range, specifically the 2D-S, which is commonly utilized for icing tunnel measurements of larger drops.

Response: Measurements with the 2D-S have already been carried out at TUBS IWT, in the size range from 10 to 1280µm (Bansmer et al. 2018). These comments were made in regard to the SLD icing conditions, where the LWC is around (0-1 0.4 g m^{-3}) with sizes often extending over 250µm.

The SLD clouds exhibit a bi-modal nature (Cober and Isac 2012) and the calibration of such a cloud in the wind tunnel requires to effectively capture the first mode (of small droplets) at 4-8µm as well as the second mode round 80-120µm. As discussed earlier the probes FCDP and FSSP have size limitations. Although shadowgraphy has no limitation on size simultaneously measuring both extremes of the SLD is highly challenging. The low LWC and broad spectrum of PSD of SLD conditions pose challenges for the individual measurement methods. Combinations like CCP, FSSP+OAP or others are to be employed. Although there are no such restrictions on PDI theoretically, the limited dynamic range and optimal selection of PMT voltage is difficult. Furthermore, the droplets in the cloud are sparse and gaining statistical confidence is more difficult than the conditions studied here.
This is briefly discussed in the revised manuscript.

References

Biter, C. J., et al., 1987: The drop-size response of the CSIRO liquid water probe. *J. Atmos. Oceanic Technol.* **4**, 359-367.
Response: Reference is now included in the manuscript.

Brenguier, Jean-Louis, et al., 1998: Improvements of droplet size distribution measurements with the Fast-FSSP (Forward Scattering Spectrometer Probe). *J. Atmos. Oceanic Technol.*, **15**, 1077-1090.
Response: Already cited, DOI link corrected

King, W. D., et al., 1985: Icing wind tunnel tests on the CSIRO liquid water probe. *J. Atmos. Oceanic Technol.*, **2**, 340-352.
Response: Already cited, DOI link corrected

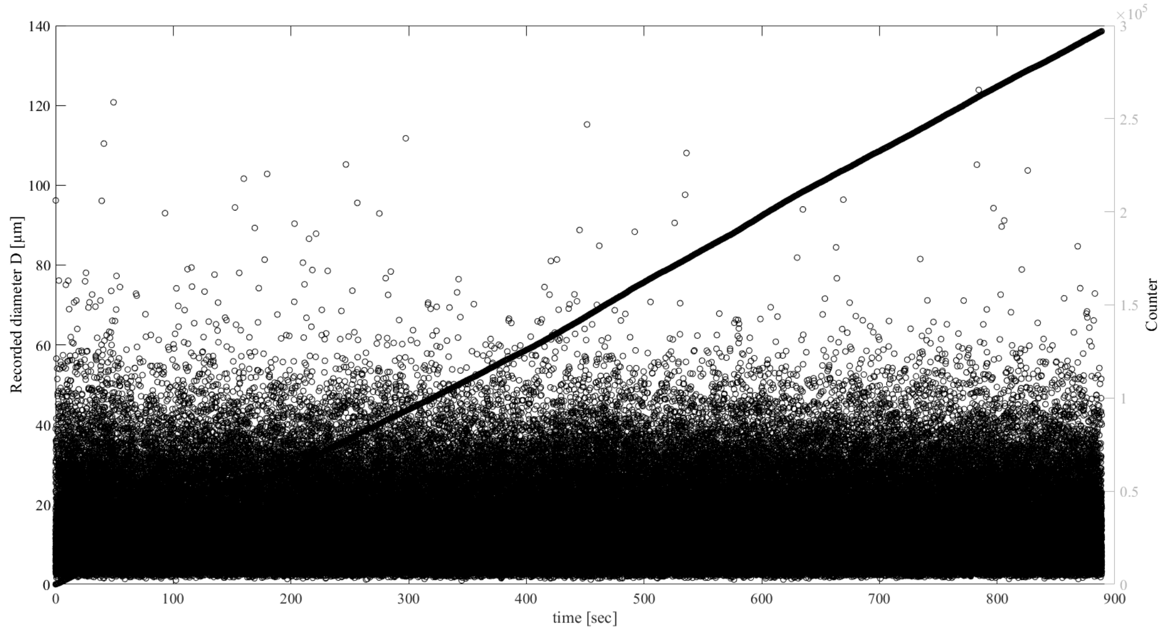
Korolev, A., et al., 2013: Modification and tests of particle probe tips to mitigate effects of ice shattering. *J. Atmos. Oceanic Technol.*, **30**, 690-708.
Response: Reference is now included in the manuscript.

Lance, S., 2012: Coincidence errors in a cloud droplet probe (CDP) and a cloud and aerosol spectrometer (CAS), and the improved performance of a modified CDP. *J. Atmos. Oceanic Technol.*, **29**, 1532-1541.
Response: Reference is now included in the manuscript.

Strapp, J. W. and R.S. Schemenauer, 1982: Calibrations of Johnson-Williams Liquid Water Content Meters in a High-Speed Icing Tunnel. *J. Appl. Meteor.*, **21**, 98–108, https://doi.org/10.1175/1520-0450(1982)021<0098:COJWLW>2.0.CO;2
Response: Cited in the introduction

Figures

Figure 2: Add mean and variance values to each plot. It would also be helpful to add time-series data for the PDI, FCDP and Shadowgraphy (counts/sec or conc/sec). Perhaps it would be better to add the probes' time-series as a new Figure.
Response: We have added mean and standard variation in Figure 4. The water flow rate exhibits strong initial transient but stabilizes approximately after 15 seconds, this results in high variance in the water flow rate. A high precision, endurance and stability of other parameters can be appreciated from the low variance.



The time evolution of the droplet acquisition of PDI is plotted above, it can be seen that the data acquisition rate is fairly linear. Further the consistent pattern suggests two minutes should be long enough for a good measurement.
This Figure is not included in the manuscript. The temporal stability of the droplet cloud measured by the PDI can be seen in Figure 6 in the manuscript.

Figure 4: Add additional accumulated size distributions for the FCDP and Shadowgraphy. If possible, overlay distributions from all three methods on the same Figure indicating Concentration, LWC and MVD for each.
Response: We added a new figure (Figure 9) where we show and compare the accumulated size distributions of FCDP, PDI and Shadograhpy for two different droplet clouds.

Figure 5: Add R2 values.
Response: We calculated the correlation coefficients R^2 between all runs and added them in the caption of the figure to maintain the clear structure of the figure.
The high R^2 values show promising repeatability of the spray.

Figure 6: Based on the results in Figure 2, the conditions (Temp, RH) are not necessarily stable for the initial portion of the sample run. As such, it is hard to separate the MVD discrepancy as a function of true fluctuations vs counting statistics. Show size distributions for the initial 1000 droplets and the final 1000 droplets.
Response: The RH is initially unstable during start of test day, but over a few tests the tunnel will be saturated and can expect a stable humidity. We plotted the data for the initial 10000 and the final 10000 droplets in Figure 6 right and no large difference is found between them (overall 280000 droplets, duration 900 sec). This shows the temporal stability of the spray and complements the data in Figure 6.

Figure 7: This plot is very useful, but as with Figure 4 it should be amended to include lines for FCDP and Shadowgraphy. It may also be helpful to interpolate the higher resolution PDI data into the FCDP size bins for a more accurate comparison. Note that the FCDP size bins are chosen to smooth out Mie bumps and to improve sampling statistics for the larges drops. Add a dashed line to indicate the threshold for rejecting data due to inadequate data points in a bin (e.g., 100 counts bin-1).

Response: we have added a comparison of the droplet counts (fig. 12) as well as the relative cumulative volume curves in figure 9 for PDI, FCDP and Shadowgraphy.

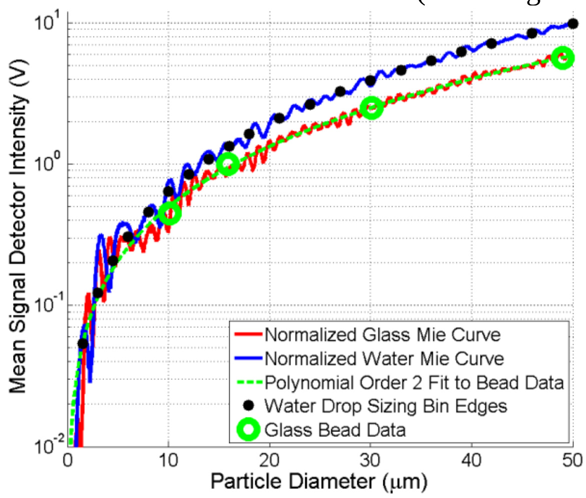
Figure 9: These plots are useful, but additional plots should be added to show correlation with LWC and Concentration.
Response: The correlation of the LWC-ratio of the FCDP to the number concentration is shown in figure 15 right. We have additionally added a figure showing the number concentration from PDI and FCDP.

Figure 11: It hard to visually separate the WFR grey region from the PDI data points. Considering switching this plot to color or making individual scatter plots for PDI, FCDP and RCT
Response: We have changed plots to color.

Figure 14: Combine with Figure 12, and include sample plots for FCDP.
Response: We have combined the two plots and changed them to color.

Tables
Table 1: Include the model number of the PDI in the caption.
Response: The table is appended with additional important parameters

Table 2: Amend the table to include these values:
FCDP Beam Waist = 80µm
FCDP DOF Rejection Criteria = 0.9
FCDP Sample Area = 0.09mm^2
FCDP Size Range = 2-50µm
FCDP Serial number = 6
FCDP Calibration Date = 4/28/2017 (see sizing calibration curve below from manufacturer)



Response: thanks for sharing the data, appended the table

Table 3: Include details on the Shadowgraphy optical system similar to Table 1 for the PDI (wavelength, magnification, focal length, working distance, collection angle, etc.).
Response: Added details of the optics

Table 4: Define the table column variables in the caption.
Response: Changed the table to give a better overview and do not further use any variable names.