We thank the Reviewers for their detailed and distinct comments towards our second revision of the submitted work. Please find our responses below.

Reviewer 1:

The Manuscript is improved with the addition of figures and descriptions suggested by both reviewers. In reading the comments from Reviewer 2, I see that she/he has carefully examined the manuscript and posed critical questions that need to be addressed. Unfortunately, in their reply, the authors have pushed back on the most critical questions from both reviewers and have only provided facile explanations.

This reviewer patently disagrees with the statements in the revised manuscript:
For the FCDP, the high sensitivity of the transit time filter to velocity differences of the droplets or a respective low sensitivity to larger particle sizes (>35 µm) was hypothesized.
Even though the qualifier, "was hypothesized" is added to the statement, there is no evidence to suggest that the FCDP has a low sensitivity to drops > 35 µm.

This statement follows two observations which have been made during data analysis:

1) The application of the TransitTime coincidence correction software provided by SPEC shows, that larger droplets along the transit time to diameter distribution are getting discarded due to a deviating fit gradient at larger droplets. Mesh points for the fit function are positioned only close to smaller diameter, where the distribution curve shows a large change in gradient. Where the so derived fit curve holds for smaller particles, a larger deviation towards increasing diameter was detected.

2) Compared to the other instruments an underestimation of the FCDP with droplet sizes larger 35µm has to be considered. Hence the phrase "was hypothesized" was used. Other recent studies (by other authors), which involve a FCDP 2D-S combination find a similar behavior in flight conditions, thus outside an artificial environment of a wind tunnel. A publication addressing this is in preparation.

The statement that conflates two inappropriate references is particularly misleading:

A hint is available is the study by Thornberry et al. (2016), where the authors only use 12 size bins (out of the 21) up to only 24 µm for data evaluation. Larger sizes are covered by a 2D-S probe with a diode array resolution of 10 µm. Sizing (and imaging) capabilities of imager probes in this size range is subject to large errors (Baumgardner et al., (2017), …). Thornberry et al. (2016) even says while comparing the size range between 24µm-36µm of FCDP and 25µm-35µm of 2D-S respectively,

"This change (projected area of measured particles by 2D-S and FCDP) *in the relationship between the FCDP and 2-D-S is due to a greater decrease in the particle concentration measured by the FCDP in the 24–36 µm size range than that measured by the 2-D-S in the 25–35 µm bin."* So the change in his linear fit over the median projected area σ in FCDP measurements is attributed to a lower number concentration of larger particles >24µm compared to what the 2D-S has observed in the given size range. But on the contrary Lawson et al. (2017) find a good agreement between the overlap region between FCDP and 2D-S.

The Thornberry et al. (2016) study was conducted in the TTL in regions with very low concentrations of ice particles (median value 18 L^{-1} according to Thornberry et al. 2016). In striking contrast, the studies in the BIWT are with water drops in concentrations of 1 to 2 x 10^6 L^{-1}. The sample volume of the FCDP is much smaller than the 2D-S for particles > 25 µm, so it is appropriate to use the 2D-S measurements for the larger particles in the TTL study. This has nothing to do with the ability of the FCDP to count drops > 25 µm in concentrations of 1 to 2 x 10^6 L^{-1}. It is only because the 2D-S has better sampling statistics at the larger (> 25 µm) particle sizes. The Thornberry et al. (2016) reference simply does not apply to data collected in the BIWT.

The statement by Baumgardner et al. (2017) is also misrepresented. Gurganus and Lawson (2018) show that the largest counting uncertainty for the 2D-S is in the 10 and 20 µm bins, and that the uncertainty is significantly reduced at the 30-µm bin and larger. Also, the Baumgardner et al. (2017) article is referring to an OAP with 25 µm pixel resolution, so the first two bins with the largest uncertainties are the 25 and 50 µm bins.

The reviewer clearly states that the FCDP's reduced sensitivity towards larger droplets would be speculative. In order to rule out a different meaning with the word "sensitivity" we also could propose: sample area size evoked reduced statistical counting efficiency, under application of the post processing parameter (DoF_{crit}). The term sensitivity is not intended to address a missizing in this droplet diameter range. In our view this effect is exactly pinpointed with an instrument's sizing efficiency or its sensitivity.

The Baumgardner et al. (2017) (fig 9-4) publication gives an example of a simulated 2D-S response to a 10 µm droplet, where (over-) sizing errors, due to out of focus transits through the laser beam result in overestimation between 110% - 190% (uncorrected) and 130% (corrected, mean), with significantly larger deviations towards farther away from the center of focus.

This is just below the regarded size bin range used by Thornberry et al. (2016) and, as pointed out by the reviewer, Gurganus and Lawson (2018) yield good agreements as low as in the 30µm size range for their lab calibration. However a newly published analysis by Oshea et al.(2020) address the presence of a small ice mode in ice cloud measurements by OAPs (in this study they refer to a CIP-15 and a 2D-S probe) might be caused by an intrinsic instrument error, immanent in this type of instrument.

We agree, concerning the Thornberry et al. (2016) publication, that the approach to prioritize a 2D-S with its larger DoF over the FCDP for cirrus measurements in the TTL, is a reasonable measure. Consequently, we no longer regard this citation as a watertight reference to support the point and refrain from using it in the manuscript.

There are several examples of good overlap between the FCDP, FFSSP and 2D-S probes in water and mixed-phase clouds. The manuscript needs to delete the hypothesis that the FCDP is less sensitive to drops > 35 µm. What the manufacturer of the FCDP does state, however, is that the FCDP erroneously sizes some drops > ~ 25 µm that are outside the DOF, and that these signals fall into the smallest size bins. This can account for the higher concentrations of drops in the ~ 5 to 8 µm size range shown in Fig. 9.

We thank the reviewer to point this out. This effect is not uncommon among forward scattering probes e.g., the Cloud and Aerosol Spectrometer (CAS) manufactured by DMT.

However, the FCDP is also felt to be more sensitive to detection of actual drops in the 5 to 8 µm size range, so this is a conundrum. On the other hand, as pointed out previously in the Chuang paper, the PDI may be less sensitive to drops in this size range. The manuscript should only discuss these sizing anomalies and not jump to the conclusion that the FCDP is less sensitive to drops in the larger size bins.

The possible error in the PDI measurements due to its smaller sensitivity to small droplets is mentioned in line 479 with the reference to the publication of Chuang.

It should also be highlighted that the FCDP (i.e., FAST cloud droplet probe), is designed for applications on research aircraft that fly at airspeeds from about 100 to 200 m s^{-1}, not in an icing tunnel at 40 ms^{-1}.

Although this aspect already has been addressed in the present manuscript version, this consideration will be further highlighted. Nevertheless, the manufacturer SPEC Inc. certifies the capability of particle sizing within a velocity range between 10 -200 ms^{-1}.

My main concern, however, is with the implication throughout the manuscript that the BWIT can be used for studies of aircraft icing. The seventh and eighth words in the Introduction are "aircraft icing". While the sentence is valid, this sets the stage for less subtle inferences that the BWIT is suitable for studies of icing of aircraft with airspeeds of 100 m s^{-1} and faster, which includes all commuter and transport class aircraft.

We made no such statement of "icing of aircraft with airspeeds of 100 m s-^{1} and faster" in our manuscript. For us, it is clear that the Braunschweig Icing Wind tunnel is not intended to use for aircraft certification. It is a research tunnel, with a certain size, speed and operational envelope, which is specified in the manuscript and also in the references mentioned. Indeed, the BIWT is used for many research topics, including aircraft icing. We are looking back on a track record of nearly 10 years of collaboration with airframers. For many investigations, it turned out, that it is not mandatory to go up to the flight speed with the investigations.

A much more egregious statement is found in the Summary where the manuscript suggests that the BIWT can be equipped for studies of SLDs in accordance with Appendix O. This is simply not true because aircraft that operate in icing conditions typically fly at 3 to 5 times faster than the maximum tunnel velocity.

We made the statement "future plans are to further enhance the capacity of the Braunschweig Icing Wind Tunnel's spray system to generate bimodal droplet size distributions according to EASA CS 25 Appendix O." We do not say anything about tunnel speed. Our approach of upgrading the tunnel with the mentioned capabilities is supervised by airframers their TIER-1 suppliers. We trust their experience and judgement.

The manuscript makes the argument that similitude (Anderson 2004) can be used to configure the BIWT for studies at velocities higher than its maximum velocity of 40 m s^{-1}. This does not appear to be a valid argument. The principal dimensionless numbers used in icing studies are the Reynolds number (*Re*) and Weber number (*We*). *Re* is the ratio of inertial to viscous forces and is proportional to airspeed. A typical chord width of an airfoil that can be tested in the BWIT appears to be about 50 mm, which is roughly the cross section of the FCDP. *Re* for an airfoil that is 50 mm in width at an airflow of 40 m s^{-1} at -10 °C is 1.6 x 10^5, which is considerably less than 2 x 10^6 reported in the manuscript. An explanation of the calculation in the manuscript should be provided. *We* is the ratio of kinetic energy to surface tension and is proportional to airspeed squared. *Re* largely controls the flow around an airfoil and *We* influences the shapes of accreted ice (along with the Nusselt number and other less significant factors). In the figure below, Anderson (2004) shows that to achieve similitude for a reference velocity (*VR*) of 200 mph (89 m s^{-1}), the scale (tunnel) velocity would have to be 400 mph (179 m s^{-1}) for a 4:1 ratio of actual airfoil chord to an airfoil in the tunnel. That is, the smaller the airfoil used in the tunnel, the larger the tunnel velocity has to be to achieve similitude. Or in other words, at the maximum velocity of the BIWT tunnel (40 m s^{-1}), the size of the airfoil would be much larger than the BIWT.

Our exact words are “To further exceed the operational envelope of the tunnel, numerous scaling methods based on similitude of geometry, droplet trajectories and the impingement heat transfer are available.”, we then further say, “In the present study, we do not apply any scaling to the results in order to avoid introducing additional sources of uncertainty to our results.”.

Since in the present manuscript, we are investigating the comparison of different droplet measurement techniques, and not aircraft icing, we should not engage into a discussion of scaling laws of aircraft icing. Nevertheless, we want to give some clarifications on the comments of the reviewer.

Typically, the chord lengths of airfoils that are investigated in icing wind tunnels are much larger than those you would apply in classical aerodynamic testing. One example can be found for instance on the homepage of the NASA Icing Research Tunnel (https://www1.grc.nasa.gov/facilities/irt/, accessed on Jan 15, 2021). The reason for that is that icing researchers want to achieve similitude at those locations, where ice accretion happens, at the leading edge.

It is misleading from the reviewer to state the Reynolds number and the Weber number are the principle dimensionless numbers of icing studies. Because the scaling of icing studies is such a difficult task, many more parameters need to be considered, e.g. the stagnation line freezing fraction and the accumulation parameter.

Obviously, the BIWT cannot be appropriately used to study inflight aircraft icing, but may be suitable for icing on wind turbine blades, slow flying UAVs and power lines. This needs to be explicitly stated in the manuscript, because as it is currently written, it appears that the reference to Anderson (2004) implies that a similitude approach can be used to study icing on aircraft wings.

The BIWT was appropriately used for several studies with commercial airframers in the past. We also refer to the above argumentation.

Also, the implication that the BIWT is in the same category as the Glenn and Ottawa high-speed icing tunnels is misleading.

We made no statement of "categories" in our manuscript. The reviewers brought up the icing wind tunnel facilities at NRC in the previous review, and therefore we mentioned them in the manuscript, highlighting the excellent collaboration with the researchers at NRC. Indeed, there is a high appreciation between the icing researchers at NRC, NASA and those in Europe, including their complementary experimental facilities.

Reviewer 2:

Most of the reviewer's comments have been taken into account and the authors did provide comprehensive answers to my questions. I think both the article and the response to reviewer include relevant information and interesting discussions for everyone in the field. Therefore, it is my belief that this article is worth being published in AMT.

Minor/technical correction:

section 2.1: the reader is referred to section 4.1 for an analysis of Fig. 2 but I couldn't find any discussion of Fig. 2 in the section 4.1.

We now included a short discussion of Fig. 2 in section 2.1 and deleted the reference to section 4.1.

section 2.2: (suggestion) depending on the final text formatting, it is sometime helpful to insert a reference to an equation in the sentence where it is defined, such as "The repeatability of measurements is characterized based on the coefficient of variation (σ) i.e., the standard deviations over several repeated measurements (n) normalized by the mean values as in eq. (2)"

We included at some positions the proposed references to equations and added also the used symbols.

section 3.2: (typo) MWD-> MVD (typo)

Corrected.

fig. 9 and discussion related to it in section 4: the "PDIFCDP" notation is unclear: is it PSD measured by the PDI during PDI vs FCDP tests ?

We now included an explanation at the beginning of section 4.