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. 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.

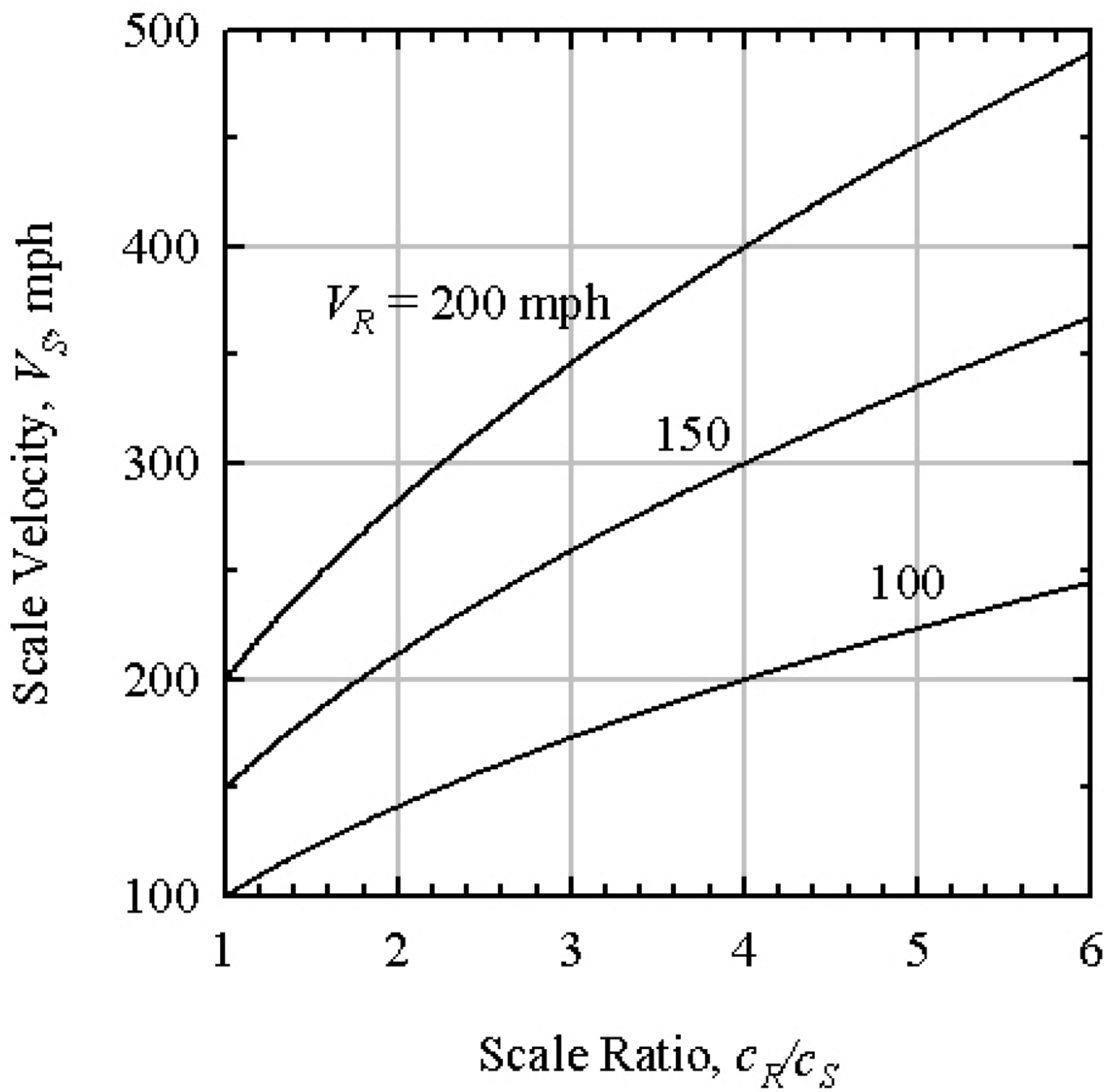
The Thornberry et al. (2016) study was conducted in the TTL in regions with very low concentrations of ice particles (median value 18 L^{-1} according to Thornberry et al. 2016). In striking contrast, the studies in the BIWT are with water drops in concentrations of 1 to 2 x 10^6 L^{-1}. The sample volume 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.

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. 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. 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 m s^{-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. 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.

The manuscript makes the argument that similitude (Anderson 2004) can be used to configure the BIWTfor 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 (V_R) 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. 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. Also, the implication that the BIWT is in the same category as the Glenn and Ottawa high-speed icing tunnels is misleading.



The manuscript represents a considerable amount of work, and the authors are commended for their efforts. However, my recommendation is that this paper be retained as a Discussion paper.