General comments:

This technical paper reports on a wind tunnel experiment designed to compare several droplet measurement techniques. The experiment is conducted in the Braunschweig Icing Wind Tunnel where populations of supercooled droplets with size ranging from 1 to 150 µm are generated. The analysis focuses on Median Volumetric Diameter and Liquid Water Content, two key microphysical properties in the characterization of icing conditions.

The droplet measurement techniques involved, namely Phase Doppler Interferometry, shadowgraphy and FCDP, a commercial single particle counter, are commonly used by cloud physics and icing research groups. Nevertheless, there are still gaps in the understanding on their respective performances which is detrimental to the comparison of data produced by various research groups using different instruments. According to the authors, wind tunnel experiments offer a unique opportunity to test droplet measurement techniques in controlled and repeatable test conditions which in the end contributes to the definition of measurement standards. Thus, the present study may bring some valuable contributions to the field and deserve a publication in AMT journal.

However, as highlighted in Kapulla et al. 2007, a thorough interpretation of the experimental results is necessary to draw a fair comparison between techniques based on different sizing and counting principles. In my opinion, this paper needs further elaboration regarding the presentation of the test conditions and the analysis of the data:

- The article does not contain a test matrix summarizing the experiment and providing the following information: wind tunnel settings (temperature, LWC, airspeed …), number of runs for each test conditions, duration and number of points collected by each probe in each run. The information scattered in the article indicates that several directions have been investigated (e.g. measurements in various wind tunnel conditions or influence of some probe settings) and that the analysis is based on a substantial number of data points, but it is hard to identify clearly the scope of this experiment and the statistical soundness of its results.
- MVD and LWC are inferred from particle size distributions (PSD) in all but two cases (rotating cylinder and tunnel air and water flow supply system settings). Given the importance of the measured PSD, the analysis given in section 4 should include a discussion on the PSD measured by the aforementioned techniques in the same test conditions. This would provide a solid basis for the subsequent interpretation of MVD and LWC results.

I strongly encourage the authors to deepen the analysis in order to strengthen their conclusions.

You will find below a detailed list of my observations and questions.

<u>Specific comments:</u>

Abstract:

(general comment): could you state the range of conditions in which the presented results apply (at least a range of LWC and MVD and the type of shape/model characterizing the droplet size distributions generated at BIWT)

l17-18: about the agreement of 15 % in MVD: the validity range of this results should be indicated. For instance, regarding shadowgraphy, your experiment shows that the indicated 15% are only valid for MVD < 35µm, see discussion in section 4.

l21-22: (question) is it an agreement between the two techniques or an agreement of each of these techniques with the reference values calculated from the mass flow rate ? In the first case, the result should be discussed in the paper in order to be included in the abstract. In the second case, the conclusion need to be rephrased, because it seems to contradict the results presented in fig. 11, on which a significant number of the PDI values fall outside the ±20% cone. From discussion in section 4.3, LWC from the PDI may only be within 20% of the reference values in 65% of the cases (97 out of 280 test points, as estimated from the data provided in section 4.3) or fall into 1:1 correlation with ±43% (whatever this means) in 91% of the cases.

Section 2:

(General comment): Add in this section a comprehensive description of your experiment. It could be test matrices summarizing the test points in terms of W/T settings and environmental conditions, targeted MVD/LWC values, number of runs, duration and number of measurement points for each instrument.

(Suggestion): To facilitate its readability, the section could be subdivided into three paragraphs: 2.1 Description of the experimental setup (already existing), 2.2 Presentation of the test conditions (new, test matrix) and 2.3 Assessment of repeatability (group together all the already existing pieces of information mentioned throughout the paper)

l109: (suggestion) provide the fluctuation level (0.1 bar) in relative units.

l127-128: “Here, we indicate the distributions and their fits in the respective experiment”: really good idea, but this has not been done, unfortunately.

l146: describe the test points and indicate the number of repetitions for each test point (use a test matrix for instance)

l147: regarding the repeatability: how do you calculate the “standard variation” for PSD? Is the standard variation equivalent to the coefficient of variation defined in equation 2?

l149: standard deviation (in g/m3) or coefficient of variation (in %)?
(suggestion) : as a complement to the comment l146, you could add a recap table (test matrix + table of results) containing test conditions, number of repetitions and statistical results (mean and standard deviation).

l159: include the test matrix here or in appendix. This is essential to give a comprehensive representation of the physical and statistical basis supporting this comparative study.

l165: This might be really interesting for your instrument assessment, since the measurement results might depend on particular instrument settings. State for each instrument, what parameters were varied and what are the results and conclusions?

Section 3:

l172: (general comment): can you indicate the general specifications of this instrument: size range (is it the static range in table 1?), velocity range, concentration range?

l186: in table 1: can you indicate the two setups (manufacturer settings, McDonell and Samuelsen 1990) in two different columns for the sake of clarity?

l191: (question) Is 5 % related to the differences obtained by repeating the tests with different user-controlled settings or is it just an indication of the repeatability of the PDI technique with McDonell and Samuelsen 1990 settings?

l192: D_{32} is not defined. Is it comparable to MVD? What point do you intend to make by quoting the results of McDonell and Samuelsen 1994?

Equations (4) to (7): please make sure that each term in equations is properly defined (e.g. what is j in $t_{tran(i,j)}$?), so that your article is self-consistent.

l200: The reference Zhu 1993 is not in the reference list

l228: The FCDP is not used in the experiment reported by Voigt et al. 2017. Please remove this reference.

l233: Do you use data of the 21st bin (over-size bin) in your calculation? Do you use a binning different than the default one set by the manufacturer?

l235-236: Regarding uncertainty: when dealing with the FCDP, you assume implicitly that FCDP, CDP and even FSSP are truly equivalent, so that you can take conclusions derived from studies on FSSP/CDP as granted for FCDP. Although all these probes use the same measurement principle (forward light scattering) and may share a similar optical layout, they differ in many aspects (e.g. the "novel fast electronics" highlighted l259). Can you provide references to studies demonstrating clearly the strict equivalence between FCDP and CDP/FSSP?
If there are no such references available, please make it clear when you discuss uncertainty that you are referring to studies on CDP/FSSP probes for lack of more relevant references. Then just mention the most relevant ones.

l241: The 32-34% accuracy range reported in Baumgardner 1983 is likely not applicable to your study ("old" FSSP with limited electronics).
l243: I think the CDP tested in Lance et al. 2010 differs from the FCDP you use handle coincidence quite differently.

l247: The content of table 2 and/or the description of the measurement protocol has to be expanded (see App. C in Lawson et al. 2017) based on your own data processing settings (e.g. what is set in the setup.m file, binning options). Also, please indicate your calibration protocol.
(question): Does "DOF_crit = 0.9" mean that particles with Qual/Sig < 0.9 are discarded? How was the value (0.9) determined and did you assess the impact of this setting on MVD for instance?

l250-254: Could you be more precise in the description of the correction algorithms applied in post-processing? For instance, the 125% threshold in beam transit time is not directly mentioned in any of the three papers you quote.

(question) How do you estimate the transit time vs drop size relationship? Do you comply with the "Half peak transit times versus size" procedure proposed in the FCDP post-processing manual?

l260 (suggestion): this assertion needs to be quantified. It would make more sense to move it into section 4.

l277: can you indicate the "data rate" in table 3.

(general comment): Some of the “characteristic numbers” given in tables 1 - 3 are interesting, but it is hard to get a clear picture of the capacity of each setup due to the lack of common parameters. A recap table with comparable specifications such as size ranges, size resolution, sampled volume, concentration range, uncertainties, main characteristics and limitations … would be useful!

l310-324: (suggestion) to be move to section 2 to establish the repeatability of the test conditions.

l324: Can you provide quantitative estimates of the uncertainty in LWC derived from the wind tunnel settings (see also comment @l380)?

section 4:

l330: to support this assertion, you can either show it analytically or quote Lance et al. 2010 (best).

l332: (suggestion) I would remove this general statement drawn from Tropea 2011: it results from a broad overview of optical techniques and does not serve your work.

l334: about the title and the content of this section
(general comment) If we follow the logical construction of the paper, the repeatability of the test conditions is not a result, but a prerequisite for the comparison of measurement techniques. Since PDI is the reference method for assessing the repeatability of the test conditions, the discussion shall be moved in section 2.
(general comment): How do you define accuracy? In this study, “precision” sounds more appropriate than “accuracy”.

l340: table 4: the test conditions and number of points underlying this table are not clearly stated. For instance: how is calculated the 5% value given in the cell (2,2) ? I assume this is the mean value of an unknown series of coefficients of variation, each obtained from several repetitions made at the same test points, but it needs to be clarified (test matrix).

l341: it is a good idea to assess the impact of the instrumental settings on the measured quantity. Please provide a detailed description of the setting being tested (test matrix…) and their impact on PSD or MVD/LWC.

l342-343: Unless the change in parameter settings is insignificant, it will make more sense to discuss separately the impact of different instrumental settings and the repeatability of the measurement techniques configured with “optimal” setting.

l357: “precision” rather than “accuracy”…
(question) For FCDP: did you investigate the impact of post-processing settings (inter arrival algorithm for instance) to retrieved PSD, as you did for the PDI?

l364-365: (suggestion) Are these two references useful here ? 1) The argument is already given line 330 (Lance et al. 2010) and 2) neither Baumgardner 1983 nor Tropea 2011 are actually dealing with the FCDP.

l376-381: this should be in section 2, in which the repeatability of the test conditions is discussed. The calculation of LWC from the wind tunnel settings and its associated uncertainty shall be discussed all in the section (experimental setup).

l383-384: this assertion should be moved to section 3.1, in which the PDI measurement techniques are introduced.
(Suggestion): is the reference to Basu et al. 2018 really relevant to this discussion? You’ve already provided enough convincing references related to the PDI measurement technique, while this one redirects the reader to a book dedicated in the first place to the physics of sprays for combustion and propulsion.

l391: (question) is 14 % the largest relative difference found between FCDP and PDI MVD (marked measurement in fig 9 left) over the entire dataset (43 data points as estimated from fig 9)?

l392 : (question) why is 5 µm the lower limit for comparing FCDP and PDI spectra? From tables 1 and 2 both PDI and FCDP seems to measure below 5µm.

l396-397: “A low sensitivity of the FCDP to larger particle sizes (> 30 µm) ….the PDI for large droplets” : what makes you think that FCDP has a low sensitivity to particles larger than 30 µm? Is it a well-known behavior of the FCDP probe? If yes, could you provide references supporting this assertion? Secondly, the argued velocity deficit for large droplets is hardly convincing: on fig 10 the density looks equally spread around unity for droplets below 50 µm (as far as I can see on my grey-printed scale picture). Finally, it would be really helpful for the reader to see how the PSD measured by PDI and FCDP differ, because at this point, one could argue that a lower MVD could either be caused by an overestimation of the number of particles in the small size bins (e.g. due to shattering), or more likely an underestimation at large end of the spectra due to poor statistics in the large size bins, as it is argued l430 during the PDI-shadowgraphy results discussion. Are the MVD values calculated from PSD integrated over the 120 sec duration reported l258 ?

l400: Have you conducted a sensitivity study, where the transit time filter is changed, in order to reach this conclusion?

l403: The references “Lance et al. 2012” and “Lance et al. 2017” are missing in the reference list.

l411: When you write “this effect can have a minor…”: have you actually assessed the effect of shattering, if any? A possibility would be to count the number of particles removed by the arrival time algorithm (provided you enable it during post-processing). The Spec software package v14 (old) contains a Quality Check program allowing to plot particle counts after the noise, shattering, DOF and TT qualification filters are applied. Such an analysis would be more convincing than the cited literature.

l411: The reference Weigel et al. 2017 is not in the reference list

l413: The ice accretion shown on figure 3 is quite impressive. Is it just an extreme case shown for illustration purposes? How much time does it take for this ice accretion to build up and how close is it to the sampling volume? Could you please comment on FCDP operation in such off-design conditions: do you see variations in the measured size distributions as the ice shape grows? Do you discard data after some changes are noticed?

l416: the references to Faber et al. 2018 and Braga et al. 2017 may be misleading because neither PDI nor FCDP is included in these intercomparisons.

l422-423: Shadowgraphy instead of here? Could you double check the data

l422-423: 8 measurement points (>35 µm, 20% of your 40-point dataset according to l278) have been excluded for being consistently different (systematic underestimation) from the expected values. Discussing the discrepancy in the PDI and shadowgraphy results found for MVD above 35µm, you suggest that a technical limitation of the shadowgraphy technique makes it unable to measure PSD correctly (insufficient statistical sampling) but your main conclusions (l18 and l510) assert that MVD measured by shadowgraphy and PDI lies within 15 %. Judging from the stated R^2 coefficient, I presume that the 15% value is only applicable if the 8 data points are discarded from the analysis. Therefore a caveat should clearly state that this is only true for MVD < 35 µm.

(suggestion): your experiment reveals a practical limitation of the shadowgraphy technique (at least when configured as in your experiment): this can be a valuable information for other W/T operators using this technique. Could you comment on whether or not this limitation is only applicable to your set up (low data rates, small field of view) or whether it is general to shadowgraphy (field of view against resolution dilemma, laser flashing rate limits) and what kind of modifications could be made to improve sizing and counting of log-normally distributed droplets from 1 to 150µm (e.g.: how to improve the data rates)?

l428: Could you quantify "very low"? Data rate should be mentioned in table 3

l436-439: These general comments do not bring useful information at this point in the discussion. Suggestion to move these two sentences in section 3.

l468-469: "This can only be explained by higher particle number concentrations measured by the FCDP ": possibly yes, given that MVD from PDI and FCDP are very similar below 20µm. Please show the measured PSD for these test conditions.

l466: what is the mean absolute value of the relative error between LWC_{FCDP} and LWC_{WFR}?

l466: (general comment) It is surprising that an instrument which only detects particles over the first third of the total size distribution overestimates the LWC! According to fig 13 right, largest overestimations (of factor of two) are registered for small MVD values, in which case FCDP measurement should be in principle most accurate (particles within its measurement range). The quoted references report overestimations ranging from 20% (Faber et al. 2018) up to a factor of 4 (Rydblom et al. 2018). Your study could potentially bring new insights and precisions on this matter, provided that the analysis is deepened. The fact that the conclusions in Lance et al. 2010 are opposite to yours (l484) raises once again the question: how far should CDP and FCDP probes be considered equivalent? If probes are truly comparable, why do your study reaches the opposite conclusions?

section 5:

l506: (suggestion) "test" instead of "boundary" conditions?

l509-510: The statement that the shadowgraphy values fall with 15% needs to be rephrased (range of validity, caveat about the low sampling rates, resolution vs sampling volume).
l512: "For the FCDP, the high sensitivity … (>35 µm) was **hypothesized**", rather than determined, since the discussion in its current state is hardly conclusive.

l515: (suggestion) this is an important conclusion but could you rephrase this, so that the limitation of your shadowgraphy setup appears clearly (low sampling rate more likely) and if possible, provide some piece of advice to others on how to improve the performance of the shadowgraphy technique in such test conditions.

l521: quantify "significantly".

page 29

Fig 10: Can you quantitatively comment the hypothesis made l95 about the drop velocity with respect to the air speed based on the PDI data collected in various test conditions?

<u>Technical corrections</u> (compact listing of purely technical corrections, typing errors)

l52: Similarly, to the experiments conducted here, Ide (1999) compared: first comma to be deleted

l345: is the reference to section 4.1 correct or should it be section 3.1?

l401: reference instead of reverence

l476: LWC_{WFR} rather than LWC_{PDI}. This typo error prompt me to ask whether or not "PDI" was meant l468 since the comparison is made with WFR in the first place?

l498: **t**he instead of The

l505/506: a good repeatability **of/?** the MVD… (word missing)