Response to the reviewer 2 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 insightful constructive comments that have definitively improved the effectiveness and quality of the manuscript. Please find the brief summary of responses below. The relevant improvements are colored in red in the manuscripts.

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.

Response: A summary table of minimum and maximum test conditions for all the measurements is added in the manuscript and the total test matrix is added as supplemental material.

- 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.

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.

In case of the FCDP microphysical properties of a spray such as MVD and LWC are higher order products derived from a sample statistics of droplet number and size. Uncertainties in the underlying measured parameter propagate, in the case of the LWC, with the order of three.

By constraining the FCDP's considered SA as a measure to constrain the measured droplet number and such the tendency for coincidence, decreases especially the counting statistics for larger droplets.

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 and the PDI measurements give in most of the cases nearly the same number densities. The new plots are discussed in the manuscript.

I strongly encourage the authors to deepen the analysis in order to strengthen their conclusions.
Response: We strived to deepen the analysis and some of the new items in the revision are below.

- The droplet spectrum, count and the droplet concentrations of the measurement systems are compared on a finer and coarser spray (Section 4.1).
- The repeatability of the spray is quantified and the temporal stability over small samples is presented (Section 2.3)
- An attempt is made to quantify the LWC uncertainty by analogies from a similar system. (Section 2.3)

Many changes are made to improve the readability and make the paper self-consistent.

Specific comments:
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)
Response: The general valid range of MVD and LWC for this study in now listed in the abstract. Regarding the cloud distribution there is no specific regulatory requirement for the droplet size distribution of these fine sprays. Typical droplet size distributions of the BIWT are shown in fig. 3 and fig. 5. These can best be described by a Rossin-Rammler distribution. However, the description and investigation of different distribution functions is not the goal of this study.

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.
Response: The stated agreement between these three measurements is for the range MVD=8-35µm. This is now included in the abstract. Due to the maximum detectable diameter of the FCDP of 50µm no measurements are discussed beyond this. The reasons for deviation of the shadowgraphy are discussed.

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.
Response: This is an agreement of each of these techniques with the reference values calculated from the mass flow rate. Accordingly, the quantities $|E_{PDI-WFR}|$ and $|E_{rotCyl-WFR}|$ are presented in the manuscript. The deviation is found to be higher at MVD of 35µm. The sentences in the abstract have been changed to correctly summarize the results of the LWC measurements.

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.

Response: A summary table of tests maximum and minimum test conditions for the measurements is added in the manuscript and the total test matrix is added as supplemental material.

(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)

Response: Certainly, this improves the readability. The section is subdivided as follows:
2.1 Wind tunnel description
2.2 Parameters and Statistical Quantities for Comparison
2.3 Wind tunnel repeatability and uncertainty estimations
2.4 Test matrix

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

Response: The typical range of operating pressures is 2 to 5 bars, the 0,1bar fluctuation level is provided in %.

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

Response: The description of the distribution and their possible fits would open a new chapter that is out of the scope of the actual paper. The misleading sentence was deleted.

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

Response: There are several repetitions of the tests, 3 example cases with each of them with 3 repetitions are presented in figure 5. The test conditions of all compared tests can be found in the test matrix that will be uploaded as supplement material.

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 %)?

Response: The sentences have been corrected. It is the coefficient of variation of the MVD. We have additionally added R^2 values for the cases shown in figure 5.

(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.

Response: We have added a summary table of the conducted experiments in the paper and will upload the full test matrix as a supplement.

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?
Response: the discussion of the different parameter settings of each technique are not part of the paper. Our focus is on the comparison of the measurement techniques not on the investigation of every single system. We forward the reader to a lot of literature where these investigations can be found.

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?
Response: We added some additional essential specifications of the instruments in the tables 1-3 and the comparable specifications in table 4.

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?
Response: the two setups used, differ in the focal length of the transmitter and the dependent variables fringe spacing and beam waist at probe volume, that are mentioned in table 1 with a backslash.

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?
Response: the 5% value is the one obtained by McDonell and Samuelsen by their tests of the PDI sensitivity to user-controlled settings.

l192: D32 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 ttran(i,j)?), so that your article is self-consistent.
Response D32 is the Sauter mean Diameter: a representative number for the ratio of the volume to the surface area, often used in industrial spray applications. We decided to delete the sentence with the results of McDonell and Samuelsen 1994 since it is treating the D32 that we have not used in our study. We further added the definition of all terms used in our equations.

l200: The reference Zhu 1993 is not in the reference list
Response: There was a mistake in the year, we have corrected it

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

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?
Response: Here we quote the overall measuring capabilities specified by the manufacturer. For a data evaluation and further analysis we excluded the over-size bin. Information from this bin has been qualitatively recognized as a hint for the amount of droplets sensed beyond the actual size range. A droplet size calibration has been performed for the FCDP using borosilicate and soda lime glass beads. It was decided to stick with the manufacturer's bin setting from the probe checkout protocol, in order to allow for an impartialed bin assignment.

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.
Response: We have to admit, that a more obvious distinction between the mentioned probes is favourable. To reference the FCDP to an FSSP and CDP might be misleading, without clearly stating its advantages. Due to a lack of a distinct study which directly compares CDP and FCDP, these probes specifics from FSSP and CDP have been employed.
The revised manuscript will point this out more clearly.

l241: The 32-34% accuracy range reported in Baumgardner 1983 is likely not applicable to your study (“old”FSSP with limited electronics).
Response: There is hardly any publication out there which explicitly gives an accuracy for forward scatter probes. The reviewer is totally right with her/his hint, that the quoted accuracy applies for early generation forward scatter probes.

l243: I think the CDP tested in Lance et al. 2010 differs from the FCDP you use handle coincidence quite differently.
Response: again it has to be brought to attention, that the quoted values only hold for a CDP, with an older optics and electronics.

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?
Response:

FCDP SN	SN06
Calibration	as of 4/28/2017
DoF criteria:	Qual/Sig Ratio>= 0.9
SA	0.09mm^2
Transit Time method	SPEC integrated Gaussian technique
Shattered particle filter	Arrival time algorithm

Operators manual are available on www.specinc.com/downloads. Matlab Software package FCDP_SP3C_V40 has been used for processing of raw files.

By selecting a Depth of Field criterion of Qual/Sig Ratio >=0.9 all droplet scatter events which do not meet this criterion are discarded. The size of the SA where droplets fulfil this respective criterion of >=0.9 has been determined within the scope of the probe calibration via a sensitivity area map using a droplet generator (Lance et al. 2010, Faber et al.,2018). A spatial resolution of this precision mapping has been 0.25mm along the laser beam direction and 0.03mm across the laser beam. Recorded particle by particle files that come with the newer electronics implemented in the FCDP, in contrast to the CDP, allows for a subsequent assignment of SA and DoFcrit pairs during post processing.
The realization of high droplet number concentrations and the increased possibility of coincidence urges the use of a high DoF ratio in order to target coincidence. The calibration specifies a DoF ratio of 0.9 as the peak value for this FCDP. SPEC recommends high DoF ratios also for accurate particle sizing.

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?
Response: The initial step in order to reduce coincidence in high droplet number concentrations is to sharpen the DoF criterion. An additional filtering method to further reduces the influence of coincidence. SPEC provides a software module in Matlab (Vers10) with which the theoretical full peak transit time (TT) through a gaussian beam profile, depending on the droplet size and TAS can be fitted to the observed TT to size distribution from the measurement using the two fit parameter C1 and C3. Qualified scatter events that are outside the acceptance range, which is a deviation of more than 25% from this TT to size curve are regarded as coincident and are such discarded (SPEC inc. C1C3_V4 manual).

$$TT = \frac{2}{TAS}\sqrt{\mathrm{C1} * \log(\mathrm{D}^2) + \mathrm{C3}}$$

l260 (suggestion): this assertion needs to be quantified. It would make more sense to move it into section 4.
Response: Will be moved and discussed in section 4 as suggested

l277: can you indicate the "data rate" in table 3.
Response: The speed is essentially limited by the response of the camera in a single frame operation mode. The acquisition rate was 2.33 images per second. This is now mentioned in Shadowgraphy description in Chapter 3.3.

(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!
Response: We have added a table with the mean characteristics, together with the summary table of the conducted experiments.

l310-324: (suggestion) to be move to section 2 to establish the repeatability of the test conditions.
Response: We believe as WFR measurements itself is another measurement moving it will disturb the cohesion

l324: Can you provide quantitative estimates of the uncertainty in LWC derived from the wind tunnel settings (see also comment @l380)?
Response: The atomization physics of internal mixing nozzles are highly dependent 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 droplet sizes and the LWC, to some extent complemented by the uncertainties introduced by the wind tunnel performance.
The primary objective of this exercise is the probe inter-comparison for which the prerequisite is the repeatability or the reproducibility and the temporal stability. Accordingly, an emphasis is made on the repeatability of the wind tunnel and spray and an attempt is made to determine the uncertainty in LWC.

Firstly, 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 according to 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.

Secondly, the repeatability of the spray can be appreciated from the plots in Figure 5.

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 14-16 beyond that value can be attributed to the uncertainties of the individual measurement techniques.
We will mention these uncertainty considerations in Section 2.3 in the revised manuscript.

section 4:
l330: to support this assertion, you can either show it analytically or quote Lance et al. 2010 (best).
Response: We have referred in the revised manuscript to Lance et al. 2010.

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.
Response: We have removed the statement.

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.
Response: we agree with the comment and have shifted the investigations of the precision to the according sections in chapter 2.

(general comment): How do you define accuracy? In this study, “precision” sounds more appropriate than “accuracy”.
Response: Being a relative comparison with unaccounted biases it is appropriate to use precision instead of accuracy, changes are made wherever necessary.

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).
Response: The values of Table 4 are the mean values of the coefficients of variations, obtained from several repetitive measurements. This explanation is included at the beginning of chapter 3. The test matrixes of the repetitive measurements are also added as supplemental material.

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.
Response: the discussion of the different parameter settings of each technique are not part of the paper. Our focus is on the comparison of the measurement techniques not on the investigation of every single system. We forward the reader to a lot of literature where these investigations can be found.
A summary test matrix is now included in the manuscript and the detailed test matrix uploaded as supplemental material.

l357: “precision” rather than “accuracy”…
Response: has been replaced into “precision”

(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?
Response: We have post processed the data under consideration of various filter techniques available in the Matlab postprocessing routine and assessed their influence on droplet number.
The inter arrival algorithm for instance was applied within the scope of a shattering filter. This filter has been applied although droplet number before and after was insensitive towards this inter arrival filter, which supports the conclusion that (maybe also caused by the presence of the anti shattering tips and with rather small droplets and no ice crystals) that shattering had no major role.
Additionally a variation of DoF criterion has been performed with 0.7 and mostly 0.8 and eventually 0.9.

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.
Response: We agree, that this is a repetition of hinting towards the nature an error in LWC, when deriving it from droplet number and size. The subsequent references can be omitted.

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).
Response: we agree with the comment and have shifted the investigations of the precision to the according sections in chapter 2.

l383-384: this assertion should be moved to section 3.1, in which the PDI measurement techniques are introduced.
Response: This has been moved and the advantages that make PDI more robust are discussed in section 3.1.

(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.
Response: the mentioned reference has been deleted.

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)?
Response: The maximum difference in MVD is 14 % for all the 45 data points compared. The according sentence has been adapted in the manuscript.

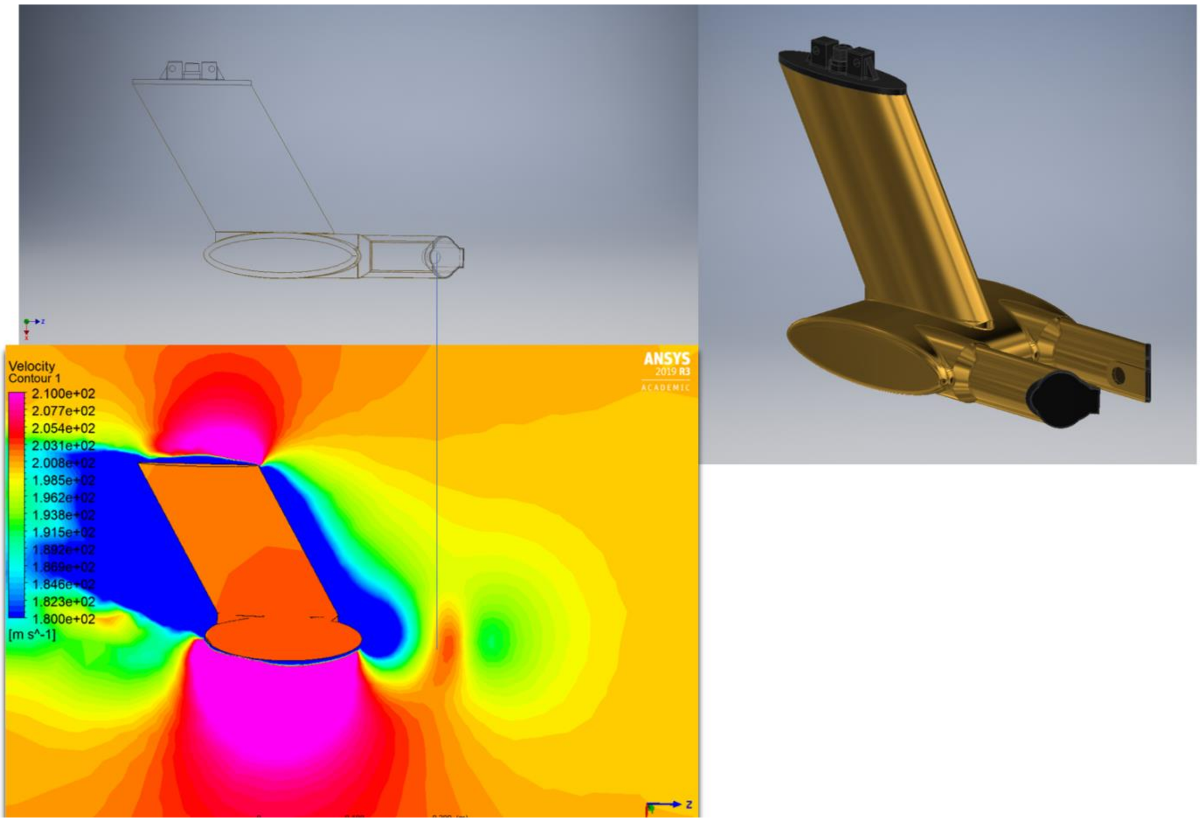
l392: (question) why is 5 µm the lower limit for comparing FCDP and PDI spectra? From tables 1 and 2 both PDI and FCDP seems to measure below 5µm.
Response: The droplet size in PDI is obtained from the linear relations between the phase shift and size derived for a predominant reflection or refraction 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 high accuracy of these small droplets but at the expense of the limiting the upper size.
A discussion of the above is included in the draft.

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?
Response: There are indeed hints of a lower sensitivity of the FCDP towards larger particles throughout various measurements, when comparing FCDP to CDP data e.g. ACTIVATE (current and ongoing NASA campaign) or in further wind tunnel tests with a FCDP, 2D-S combination at RTA, (Vienna, Austria) during the ICE GENESIS campaign. Unfortunately there is now reference available yet. 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.

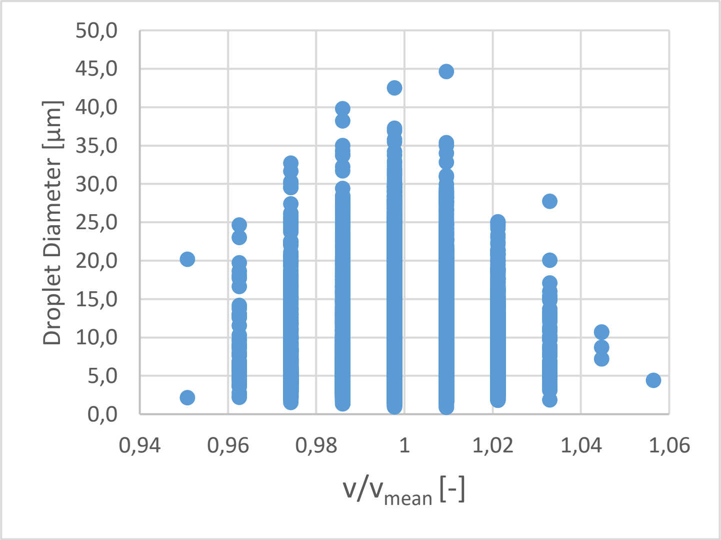
Secondly, the argued velocity deficit for large droplets is hardly convincing: on fig 10 the density looks equally spread around unity for droplets below 50 µm (as far as I can see on my grey-printed scale picture).
Response: We compare velocity measurements from the PDI, a completely non-intrusive measuring technique, with an intrusive technique. With a surface area of the test section of 50cm x 50cm and a projected surface of the FCDP of approximately 171.69 cm² almost 7% ! of the cross sectional area are occupied by the probe itself, without taking a boundary layer within the test section into account (reducing the cross section by an assumed boundary layer thickness of 1cm yields a relative FCDP cover of 8%). Without assessing the stream lines around the FCDP and droplet trajectories in detail we might have to consider this effect, especially when comparing different measurement techniques. The figure below shows the velocity field across the centre plane through the SA, around a FCDP at a given true air speed of 200 m/s. Although our measurements were conducted at a lower TAS, we want to draw

the attention towards the point that the velocity field along a potential particle trajectory ahead of the probe arms, as well as directly where the SA is located is modified by the probe itself. This fluid simulation was conducted in a free flow environment and without the constraint of a test section.
Velocity measurements with the PDI on the other hand are unobstructed and undisturbed by the probe itself.



Simulated droplet speeds for this specific wind tunnel setup vary with droplet diameter. According to this simulation this effect is pronounced for larger droplets (>100µm).
Ansys simulation results for single droplets accelerated with the 3D-airflow of the wind tunnel nozzle show for droplets >150µm a velocity deficit of 10% in the test section.

Diameter in µm	simulated velocity at test section in ms^{-1}
160	36.07
200	34.86
240	33.73
280	32.73
320	31.92
360	31.36



In addition to the particle velocity plot (figure 2) we show in the above image the PDI velocity measurement results of a test case with only small droplets. The maximal velocity deficit is less than 5%.

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.
Response (equal as above) : 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 is 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 and the PDI measurements give in most of the cases nearly the same number densities.
The new plots are discussed in the manuscript.

l400: Have you conducted a sensitivity study, where the transit time filter is changed, in order to reach this conclusion?
Response: Settings for the transit time filter have been adjusted throughout the whole data analysis process, until coming up with the current and final settings. The motivation for this proposed conclusion is the observation that the C1C3 fit routine has a good agreement along the maximum occurrence of observed transit time to droplet diameter pairs for smaller droplet sizes. Advancing to larger droplet sizes the gradient of the fitted theoretical transit time versus droplet size curve gradually deviates from observed transit times. This observation is so pronounced that larger droplets along the observed TT vs droplet diameter distribution might fall beyond the acceptance range of 125% about the fitted theoretical curve. This brought us to the proposed conclusion that particle speeds of larger droplets seem to deviate more from the theoretical TT vs diameter curve. This instance can be adjusted and was partially accounted for by manually shifting the fitted curve along the TT-axis (accepting potentially more coincident particles and allowing more larger droplets into the acceptance range). Although having observed the variation in particle speed with droplet size, this effect might not sufficiently explain the declining sensitivity with larger droplet diameter. It is more likely that the reduced sensitivity might be promoted by a statistical underrepresentation due to the strict DoF criterion and the corresponding small size of the SA. Thornberry et al.(2016) can be quoted as reference.

l403: The references “Lance et al. 2012” and “Lance et al. 2017” are missing in the reference list.
Response: Lance 2012 included, Reference “Lance et al. 2017” is a typo

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.
Response: The inter arrival algorithm for instance was applied within the scope of a shattering filter. This filter has been applied although droplet number before and after was insensitive towards this inter arrival filter, which supports the conclusion that (maybe also caused by the presence of the anti-shattering tips and with rather small droplets and no ice crystals) that shattering had no major role.

l411: The reference Weigel et al. 2017 is not in the reference list
Response: this reference will be added

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?
Response: Ice accretion in this extent as shown in figure 3 lead to an interruption of the current test point since its effect on the surrounding flow is not quantified. Furthermore probe icing of this extent also indicated icing on the flow guiding vanes of the recirculation wind tunnel. During these breaks ice build-ups have been mechanically removed and the current test point subsequently repeated. Ice-build ups of this extent have only been observed after several test points in a row, with a certain build-up time.

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.
Response: We will search for another reference in order to assess the PDI measurements. Unfortunately to our knowledge there is no single reference which juxtaposes both instruments.

l422-423: Shadowgraphy instead of here? Could you double check the data
Response: we have corrected the sentences and checked the data in the revised manuscript.

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.
Response: Yes the best linear fit of $MVD_{Shadowgraphy}=0.97 \cdot MVD_{PDI}$ is obtained by excluding the data points above 35 µm. Now this caveat is made explicitly both in abstract and conclusion.

(s
uggestion): 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)?
Response: The low data rate of Shadowgraphy setup is primarily from the camera speed and the laser. A tradeoff has to be made on the size resolution of the droplet to be captured, it should be noted that the intensity of the light source reduces with the square of the magnification factor of the teleconvertors and it leads to a point where the gradients between the background and shadow become weak and the lower thresholds specified would lead to large noise picked as smaller droplet, further the resultant area reduction also reduces the probability of the finest droplets being detected thus hampering the quality of the measurement. Higher resolution cameras and high intensity light sources will improve a better.
A more detailed description of the Shadowgraphy setup is now included in the manuscript.

l428: Could you quantify “very low”? Data rate should be mentioned in table 3
Response: The data rate was approximately 2 frames per second. This is added in the description of the setup in Section 3.3.

l430 during the PDI-shadowgraphy results discussion. Are the MVD values calculated from PSD integrated over the 120 sec duration reported l258?

Response: the 120s duration was used for the FCDP measurements, since no online direct output of the counts is available. The MVD and LWC calculation was done over a time slot with temporal constant spray conditions (starting point of the evaluation after the ramp-up of the spray system). The FCDP samples consist thereby of at least 35000 droplets and in average of approx. 60000 droplets.

All the PDI measurements are made with at least 10000 droplets per sample. This led to a probe volume corrected total counts of 20000 for individual cases with low data rates at very low LWC and in average approx. 60000 counts, independent of the duration of the measurement. The MVD and LWC calculation was done over the entire data set, since the data recording was started always appr. 20s after the start of the spray system, so the ramp-up of the droplet cloud is not included in the results.

The Shadowgraphy measurements were done for at least 15 minutes to capture a minimum of 3000 droplets. This leads to more than 10000 counts with the applied DOF and border-correction. The data recording was started always appr. 20s after the start of the spray system, so the ramp-up of the droplet cloud in not included in the results.

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.

Response: The Sentences have been moved.

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.

Response: We have added a new figure (figure 9) to compare the PSD from the three measurement techniques and its discussion in the manuscript and as well the comparison of the measured particle concentrations of PDI and FCDP.

l466: what is the mean absolute value of the relative error between LWCFCDP and LWCWFR?

Response: As per the definition used in equation 3 it is 68,2%

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?

Response: The FCDP’s overestimation in LWC is promoted by high droplet number concentrations especially measured at small droplet sizes, as can be seen in the new Figures 9. Measuring conditions in the wind tunnel lie outside the customary environment in which the FCDP normally operates. References regarding FCDP’s measuring capabilities are scarce.

The reference towards an opposite conclusion by Lance et al. (2010) was revised in the new Version of the manuscript and omitted. In detail Figure 7 in Lance et al. (2010) show a positive trend in LWC bias with increasing droplet number concentrations. Figure 15 shows simulated data where larger LWC biases are promoted by a high number concentration of small droplets rather than by larger droplets. Such a behavior indicates coincidence effects.

section 5:

l506: (suggestion) “test” instead of “boundary” conditions?

Response: It has been replaced

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).
Response: It is revised with a caveat.

l512: “For the FCDP, the high sensitivity … (>35 µm) was **hypothesized**”, rather than determined, since the discussion in its current state is hardly conclusive.
Response: It has been replaced

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.
Response: Improvements have been made in setup description and post processing description. Also some recommendation discussed previously are presented in section 3.3

l521: quantify “significantly”
Response: From figure 13 left it can be observed that LWC from FCDP varies by a factor of 0.5 to 3 of the LWC of WFR. Modified the lines to reflect the same.

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?
Response: The velocity at the test section is computed with a computational model. At 40m/s of air speed, a 100 µm will be decelerated by the drag to a velocity of 37,8 m/s (5% deficit). As the droplet size decreases, the inertia of the droplet and the drag are negligible and therefore will have the same velocity as the surrounding air all along its path from injection to the test section. Accordingly, with tunnel fluctuations (±1,5%) and measurement errors, the smaller particle are expected to have a large band of normalized velocities, the same is being reflected in PDI. Larger particles have higher inertia and little less sensitivity to instantaneous fluctuation in the tunnel and also experience considerable drag that causes velocity deficit (5% for 100 µm) this demonstrates the consistency and robustness of PDI for velocity measurement.

Technical corrections (compact listing of purely technical corrections, typing errors)
Response: All of the typing errors are fixed

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: LWCWFR rather than LWCPDI. 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)