Response to Anonymous Referee #2

(Referee comments are in black; our responses are in blue.)

Thanks very much for recognizing the value of the research, and for the thoughtful comments which helped us to improve upon the original manuscript. We've chosen to address the comments point-by-point below

My main concerns are

1) the interdependence of the thermodynamic variables with one another (they are not independent)…

Certainly the variables are not independent. This was stated in subsection 4.2 in the original manuscript, and to make the issue clearer, we added the following to subsection 3.1:

"Bear in mind that T, QV, and RH are not entirely independent. Indeed, RH is a function of both T and QV, and maximum QV (approximately equilibrium QV) is limited by T. So the three variables should not be interpreted as three separate metrics of the thermodynamic state. Rather, they represent three different related, but distinct, means of characterizing the thermodynamic state."

2) …that assumption that these thermodynamic variables explain the bulk of the cloud properties…

See below for this response.

3) …and uncertainties in the ARISE dataset that should be described in more detail.

The headers for the ARISE data files (found at https://www-air.larc.nasa.gov/) report the uncertainties of QC, T, and RH as about 5% of the measurement, 0.5°C, and 2-5% (percentage of RH, not of the measurement). An uncertainty value for QV is not provided. These uncertainties are for the un-averaged 1 hz measurements, not of the grid box averaged data. The stated cloud water uncertainty may be a bit low, as shown by Fig. 2 and the associated discussion.

Panel E shows that the disagreement between the two sensors can be more than 100% for individual grid boxes. But overall, at larger QC values, CDP tends to report (derived) QC values ~25% lower than WCM-2000. An uncertainty of this size is notable (assuming the CDP is truth), but alone is not enough to explain the 50-75% disagreement between ARISE (from the WCM) and ASR.

The manuscript text in subsections 2.1 and 3.1, and section 6 are now updated to describe these uncertainties.

Regarding the second point, I am wondering whether cloud condensation nuclei (CCN) could explain some of the large differences in the vertical cloud profiles in ASR compared to ARISE. The authors mention the lack of ice in ASR and maybe the lack of importance of ice-nucleating particles, but there are also CCN. If CCN are important, then this may imply that local sources are important and that comparisons outside the flight track may not be fair comparisons.

The analysis was based on meteorological properties that were measured by the aircraft. Of course there are other important factors controlling QC, including circulation (both boundary layer and free troposphere) and aerosols. These factors are not entirely independent. With respect to aerosols, it isn't clear how disagreements in CCN concentration would lead to a 50-75% error in ASR QC. ASR microphysics does not account for CCN (or ice forming nuclei (IFN)), and instead is set to simulate clouds in "typical" conditions (including CCN and IFN concentrations). So any aerosol-related disagreement would have to result from anomalous real-world CCN concentrations causing the Sept 2014 clouds to behave much differently than "typical".

Were there anomalous CCN concentrations over the Beaufort Sea during Sept 2014? There are a few different datasets that could help figure this out. One quick way to roughly estimate aerosol conditions during Sept 2014 is with MERRA2, where monthly mean aerosol data can be accessed via the NASA Worldview tool (https://worldview.earthdata.nasa.gov/). Looking at the "aerosol optical depth" overlay, MERRA2 does not

indicate particularly anomalous AOD conditions during Sept 2014, being greater than that of Sept 2013, but lower than that of Sept 2017, for example. Granted, this is a reanalysis product, not actual measurements, and AOD is not the same thing as CCN concentration. But it suggests that Sept 2014 *likely* did not have substantially "atypical" CCN conditions.

But what if there had been anomalous conditions? Variability in CCN concentration can change the number of cloud droplets, with droplet size varying inversely with number. This can have major effects on certain macrophysical cloud properties such as albedo, but QC? The immediate effect of CCN variability mainly redistributes QC between droplets, so there would need to be secondary mechanisms at work for anomalous CCN to significantly affect QC. It might be helpful to think of some scenarios in which CCN concentration variability can greatly alter QC via secondary mechanisms. One possibility could be that high CCN concentration in the real atmosphere during Sept 2014 suppressed precipitation rates and extended the cloud lifetimes (i.e. the Albrecht effect); if ASR did not account for this, it could produce too much precipitation and desiccate the simulated clouds. But as discussed with Referee #1, there is no clear evidence that ASR has a problem with overactive precipitation formation. Coupled with the lack of evidence for high CCN concentration, this mechanism probably isn't a major factor in our results.

So what other mechanism could there be for aerosols to play in this disagreement? Perhaps there could be some small effects by making cloud droplets more or less susceptible to evaporation from dry air entrainment (as discussed by Ackerman et al. (2004)). But any significant effect of aerosols on QC would probably have to include IFN and mixed-phase processes, not just CCN. And as stated in the paper, there does not appear to be enough QI in ARISE or ASR for mixed-phase processes to be a dominant factor. For these reasons we do not think that CCN can lead to a 50-75% disagreement between ARISE and ASR.

We added a short paragraph at the end of subsection 2.1 which briefly summarizes the main points of this discussion.

Minor comments:
1) Figures 6 and 7: It would be more informative to make those datapoints show the frequency of occurrence of the datapoints using a colour as a third dimension.

We've adjusted Figs. 6 and 7 (and Fig. 2) as suggested.

2) Lines 320 to 323: should be one sentence.

The second sentence was rewritten to be a complete sentence. Combining the two would be a bit of a run-on sentence.

3) In general, I think the writing can be clarified to state the key points more clearly. Many of the discussions can be distilled to simpler messages.

We have edited the manuscript with an eye towards to clarify the key points and distill the discussion into simpler messages.