

Response to the Editor and Reviewers

We thank the editor and both reviewers for their careful and constructive assessments of the revised manuscript. We are pleased that the reviewers consider the manuscript likely to provide a valuable resource for the ice-nucleation community. We have addressed all remaining technical, presentational, and editorial comments, as detailed below.

I (Tom Whale) would also like to apologise to the editor and AMT editorial team for being slow to respond at various points of the review process. This has been the result of various bouts of poor timing, including my extended paternity leave in the most recent case. Thank you for your ongoing patience and assistance.

Response to Reviewer 1

We thank Reviewer 1 for their positive assessment of the revised manuscript and for identifying several remaining technical and presentational issues.

Comment 1: Construction and labelling of the two-component example

Line 774: What is the reason for the difference in accuracy? The standard deviation of the first component is 2.853°C , while the same of the second component is 3°C ? The mean in this case is the temperature at which the fraction frozen curve reaches a value of 0.5? It is also not completely clear to me whether the two curves shown in Figure 9 panel a are “Curve 1” and “Curve 2” as described in the text. This should be clarified and potentially the legend in panel a should be adjusted to reflect the names as used in the text.

Response: We thank the reviewer for identifying these points requiring clarification. The differing numerical precision reflects the origins of the component parameters. The mean and standard deviation of the colder component, -26.375°C and 2.853°C , were taken directly from the fitted background-freezing distribution. The warmer component was constructed to represent a plausible additional warmer-freezing population and was assigned rounded illustrative values of -17.5°C and 3°C . The difference in precision therefore has no further statistical significance.

We have also clarified that, for each individual Gaussian component, the mean is the temperature at which that component reaches a fraction frozen of 0.5. This does not generally correspond to a fraction frozen of 0.5 for the combined mixture distribution.

Revised manuscript text:

To test this, we constructed two synthetic curves from the same two Gaussian freezing-temperature components. Figure 9a shows the underlying distributions and example synthetic experiments. The colder component had a mean of -26.375°C and a standard deviation of 2.853°C , with both values taken directly from the fitted background-freezing distribution. The warmer component was constructed to represent a plausible additional warmer-freezing population and was assigned a

mean of -17.5°C and a standard deviation of 3°C . The differing numerical precision therefore reflects the origins of the parameter values: the colder-component values were obtained from a fit, whereas the warmer-component values were specified as rounded illustrative values. For each individual Gaussian component, the mean is also the temperature at which that component reaches a fraction frozen of 0.5, although this is not generally true of the combined mixture distribution. The two curves differed only in the mixing ratio of the warmer component. Curve 1 contained 10 % of the warmer component, whereas Curve 2 contained 30 %. The two curves therefore contained the same underlying freezing components but different relative contributions from those components.

Comments on tables and figures

Figure 1

In Figure 1, the lines are described in the legend and I understand the markers to represent a freezing event. I think the visual clarity would benefit from a removal of the markers, as the step-lines already provide that information. This would also favor the highly active birch-pollen sample, where the underlying lines are not visible anymore beneath the markers.

Response: We thank the reviewer for this suggestion. We tested a version of Figure 1 without the markers, but found that this reduced rather than improved the overall clarity of the figure. The markers provide a direct visual indication of the discrete freezing events from which the empirical step curves are constructed and help distinguish the individual synthetic experiments. We have therefore retained the markers in Figure 1.

Figure 2

In Figure 2, one could think about adding a horizontal line at 0.95 to highlight the quick rise as the droplet number increases.

Response: We thank the reviewer for this helpful suggestion. We have added a horizontal reference line at 0.95 to Figure 2 to make the rapid approach to the nominal coverage level more readily visible.

Figure 4

Figure 4: The panel labels are covering parts of the y-axis labels, which should be fixed by shifting them slightly to the left. In addition, the panels should be addressed in the figure caption, i.e. in the second sentence: “(panel a) BCS375 feldspar, panel b) Betula pendula pollen washing water, and panel c) background freezing”. Also the x-axis should be shared and size of the labels needs to be increased slightly.

Response: We thank the reviewer for these suggestions. We have repositioned the panel labels in Figure 4 so that they no longer overlap the y-axis labels, increased the sizes of the axis and tick labels, and applied a shared x-axis. We have also revised the caption to identify the samples shown in the individual panels explicitly: panel (a) BCS375 feldspar, panel (b) *Betula pendula* pollen washing water, and panel (c) background freezing.

Figure 6

Figure 6: I would suggest to adjust the size of the labels, see also my comments regarding Figure 4. Figure 6 is also quite a low resolution, which should be fixed. I think it is also useful to use the same y-axis limits for all three panels, the added

benefit of a clear separation of the values for the birch-pollen sample does not add much.

Response: We thank the reviewer for these suggestions. We have increased the sizes of the axis and tick labels in Figure 6, re-exported the figure at higher resolution, and applied the same y-axis limits to all three panels to facilitate direct comparison.

Remaining technical comments

We have also made the remaining technical and editorial corrections requested by Reviewer 1:

- We corrected the punctuation in the sentence describing the behaviour of $K(T)$ at $f(T) = 0$ and as $f(T)$ approaches 1.
- We checked the relevant section, figure captions, and Supplementary Information and made the formatting of mathematical variables, including n and α , consistent.
- We standardised spacing between numbers and the percent symbol throughout the manuscript and Supplementary Information.
- We shortened and corrected the discussion of the ordinary and weighted log-rank tests. It now reads:

The ordinary log-rank test may be useful under proportional hazards or for small shifts between fraction-frozen curves, but it should not be the default for droplet-freezing data. Weighted log-rank tests are not generally recommended because there is usually no clear basis for choosing the weighting in advance.

- We checked and standardised the reference list, including publisher capitalisation, punctuation, spacing, and year formatting.

Response to Reviewer 2

We thank Reviewer 2 for their careful assessment of the revised manuscript and for recommending its acceptance. We are pleased that the reviewer considers the new statistical test and the validation of the confidence intervals likely to provide a valuable resource for researchers using droplet-freezing measurements. We have addressed the remaining minor and editorial comments as follows.

Minor comment 1: Temperature interval and grid spacing in Eqs. (19) and (22)

Equation 19: Is it likely for the choice of temperature interval and the resulting temperature grid to influence this result and therefore the p value in Equation 22? For example, if a tiny temperature interval is used such that every droplet freezing in each experiment has its own bin, would the test always indicate a significant difference between the curves, or is this not an issue because of the use of permutation statistics? On the other extreme, having only one large temperature bin would, of course, cause the test to observe no differences between the curves. It is not obvious to me from the text what, if any, quantitative differences in calculated p values would be observed with reasonable intermediate choices of ΔT (say, 0.5 versus 1 °C). Guidelines as to an appropriate choice of ΔT in this situation should be given.

Response: We thank the reviewer for raising this point. The temperature grid in Eq. (19) is used only to evaluate the integral numerically; the grid points are not treated as separate bins or independent statistical comparisons. Refining the grid therefore does not inherently increase the

likelihood of obtaining a significant result. As the grid is refined, the larger number of evaluation points is offset by the factor ΔT , and the calculated sum converges towards the integral in Eq. (17). In addition, the observed statistic and every permutation statistic are calculated using the same temperature interval and grid.

We agree that an excessively coarse grid could fail to resolve differences between the curves. We have therefore clarified the role of the grid in the revised manuscript and added practical guidance on choosing ΔT . The implementation supplied with this work uses $\Delta T = 0.01$ °C, which is small relative to the temperature scale over which typical droplet-freezing curves vary. For alternative implementations, we recommend choosing ΔT sufficiently small relative to the narrowest curve feature of interest; where necessary, numerical convergence can readily be checked by repeating the calculation using a finer grid. We have also revised the discussion of the temperature interval to make clear that it defines the portion of the curves being compared and should be specified before testing and applied identically to the observed data and all permutations.

Revised manuscript text:

The temperature grid is used only for numerical evaluation of the integral and does not divide the data into independently tested bins. Consequently, reducing ΔT does not inherently increase the likelihood of obtaining a significant result: as the grid is refined, the larger number of evaluation points is offset by the factor ΔT , and the sum converges towards the integral in Eq. (17). The same temperature interval and grid are also used for the observed statistic and for every permutation. Conversely, an excessively coarse grid could fail to resolve differences between the curves. The implementation supplied with this work uses $\Delta T = 0.01$ °C, which is small relative to the temperature scale over which typical droplet-freezing curves vary. For alternative implementations, we recommend choosing ΔT sufficiently small relative to the narrowest curve feature of interest; numerical convergence can readily be checked by repeating the calculation with a finer grid.

We also replaced the previous statement concerning the selected temperature interval with the following:

The absolute value of Q_f depends on the selected temperature interval because the statistic has units of temperature. The interval defines the portion of the curves being compared and should therefore be specified before testing and applied identically to the observed data and all permutations.

Editorial comment 2: Repetition in the abstract

Lines 19 and 29 are repetitive. I recommend removing “while recognizing that a significant global test does not identify where the curves differ or why” on line 19, and keeping line 29.

Response: We agree and have removed the earlier occurrence of this statement while retaining the fuller explanation later in the abstract.

Editorial comment 3: Grammatical correction at line 577

Line 577: test $\rightarrow$ tests.

Response: Corrected as suggested.

Editorial comment 4: Figure cross-reference at line 627

Line 627: I think this should refer to Figure 7b.

Response: We agree and have corrected the cross-reference from Figure 7a to Figure 7b.

Editorial comment 5: Repeated caveat at lines 735–736

Lines 735–736: This has been emphasized twice now in the manuscript. I think repeating the caveat a third time here is not needed.

Response: We agree and have removed this repeated caveat. The earlier explanations have been retained.

Editorial comment 6: Harrison et al. and Polen et al. references

In the references, the Harrison et al. reference is cited twice, once as the ACP Discussions paper and once as the final ACP paper, and the Polen et al. reference refers only to the preprint version in AMT Discussions.

Response: We have removed the duplicate discussion-version reference for Harrison et al. and retained the final published article in *Atmospheric Chemistry and Physics*. We have also updated the associated in-text citation as necessary. The Polen et al. reference has been replaced with the final published article in *Atmospheric Measurement Techniques*.

We thank the editor and reviewers again for their helpful comments. We believe that the resulting clarifications, technical revisions, and editorial corrections have improved the manuscript.