https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

Survival analysis for droplet-freezing data: Kaplan–Meier confidence intervals and log-rank tests

Thomas F. Whale[1], Sarah L. Barr[1,2] and Trystan Surawy-Stepaney[1]

[1]Institute for Climate and Atmospheric Science, University of Leeds, Leeds, LS2 9JT, UK
[2]National Centre for Atmospheric Science, Leeds, LS2 9PH, UK

Correspondence to: Thomas F. Whale (t.f.whale@leeds.ac.uk)

Abstract

Droplet-freezing assays underpin immersion-mode ice-nucleation research yet approaches to uncertainty quantification for fraction-frozen curves and derived active-site densities ($n_s(T)$) are inconsistent. Further, there is not currently a rigorous method for significance testing the difference between fraction frozen curves. To address these issues, we recast droplet-freezing measurements as survival data and apply analysis techniques typically used in medical statistics. Using the Kaplan–Meier estimator, we derive nonparametric confidence intervals for droplet fraction frozen and $n_s(T)$ without binning or model assumptions, matching Monte-Carlo and studentized-bootstrapped intervals on a literature volcanic ash ice nucleation dataset. Confidence intervals calculated for simulated datasets show precision improves with sample size and with steeper fraction frozen curves. Adapting the log-rank test, we introduce a method for comparing fraction frozen curves and demonstrate its application to literature and simulated droplet freezing datasets. We recommend reporting Kaplan–Meier confidence intervals on droplet freezing datasets and employing the log-rank test when comparing droplet fraction frozen curves.

1 Introduction

Droplet freezing experiments are widely used for the study of immersion mode ice nucleation e.g. (Polen et al., 2018; Whale et al., 2015; Bigg, 1953; Vali, 1971; Vali et al., 2015), a process of critical importance for the glaciation of mixed-phase clouds (Knopf and Alpert, 2023; Herbert et al., 2025). A droplet freezing experiment cools multiple samples of water, usually prepared in a physically identical manner, and measures the temperatures at which they freeze. The unprocessed data produced by these experiments are typically referred to as ‘fraction frozen curves’ or similar. Experiments of this type have been employed since at least the 1950’s e.g. (Bigg, 1953) with droplets ranging in size from picolitres to millilitres (Miller et al., 2021; Murray et al., 2012). Despite this long period of usage, there is no consensus on how the data produced by droplet freezing assays should be analysed. A particular problem is the calculation of confidence intervals or bands for fraction frozen curves, and for quantities derived from fraction frozen curves, such as nucleation rates and deterministic nucleus spectra (Fahy et al., 2022). The purpose of calculating such confidence intervals is to indicate how confident we should be in the value of a calculated datapoint. For instance, a 95% confidence interval indicates that if we repeated the measurement many times, about 95% of the intervals we compute would capture the true value.

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

Fahy et al. (2022) produced a thorough discussion of the various approaches which have been employed to calculate confidence intervals for ice nucleation data. The following briefly summarises this discussion. Where multiple identical experiments have been performed, means and standard deviations of the repeated experiments can be calculated. It is unclear how many experiments would be needed for the central limit theorem to become valid in these cases. Binomial confidence intervals have been calculated by treating fraction frozen values as binomial ratios and using an adjusted Wald interval (Agresti and Coull, 1998), however most ice nucleation data are not normally distributed, meaning this approach may not always be applicable. Poisson Fiducial limits have been used (Koop et al., 1997), which assume that variability in droplet freezing is described by the Poisson distribution. Fahy et al. (2022) state that this approach may not account for all the variables involved in ice nucleation process. The final approach uses Monte Carlo simulations to simulate experiments based on underlying assumptions about the nature of the ice nucleation process. For instance, Harrison et al. (2016) used ice-active site densities calculated from the original experiments to simulate number distributions. Under the assumption that a droplet freezes when it contains at least one active site, they were able to construct simulated experiments based on the experimental data. An alternative approach suggested by Vali (2019) and applied in various subsequent papers assumes that the number of freezing events in a given temperature interval is Poisson distributed, and generates Poisson distributed random numbers taking the experimentally observed number as the expectation value for the temperature, thereby generating simulated spectra. In both cases, confidence intervals can be calculated by generating hundreds or thousands of simulated experiments and finding quantiles of the simulated data. Fahy et al. (2022) refer to these approaches a 'parametric bootstrapping' because they require that ice nucleation measurements be parameterised before Monte-Carlo simulations are performed. This necessitates assumptions about the nature of ice nucleation related to distribution of active sites on surfaces and stochasticity or otherwise of the process, or else rely on coarse binning of data in a manner than conflicts with the exponential increase in ice nucleation rate with temperature. To avoid these difficulties Fahy et al. (2022) proposed an 'empirical bootstrapping' method for calculating confidence intervals resulting in 'studentized' confidence intervals. This method uses the observed list of droplet freezing temperatures with replacement to generate many synthetic experiments, then uses the ensemble to quantify uncertainty (e.g., confidence intervals) without assuming a physical model or a specific error distribution.

Although droplet freezing datasets can be analysed in several defensible ways, the field lacks a widely adopted, standardised statistical framework for quantitatively comparing fraction frozen curves. As a result, decisions about whether two curves represent meaningfully different INP populations are often based on visual inspection or informal criteria, which can be ambiguous when uncertainties are large or curves partially overlap. We therefore propose a simple, non-parametric method that re-casts fraction frozen curves as survival curves and uses established tools from survival analysis. This approach provides (i) convenient confidence intervals and (ii) a basis for testing curve-to-curve differences using the log-rank test, thereby reducing reliance on subjective interpretation.

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

2 Statistical background

2.1 The Kaplan–Meier estimator applied to droplet freezing data

In a typical set of droplet freezing experiments, multiple aliquots of water are exposed to a systematically decreasing temperature, allowing observation (typically visual) of the onset and distribution of freezing events across the sample set. The fraction frozen is calculated as:

$$f(T) = \frac{N_F(T)}{N_0} \tag{1}$$

Where $N_F(T)$ is the number of droplets frozen at temperature T and N_0 is the total number of droplets, frozen or unfrozen, present in the experiment (Vali et al., 2015). A typical dataset where a survival curve might be employed would have $n(t_j)$ subjects at risk of failure at time t_j and $d(t_j)$ failures at time t_j (Kalbfleisch and Prentice, 2002). The nonparametric likelihood estimate of the survivor function, the probability that a sample survives longer than time t, is then

$$\hat{S}(t) = \prod_{t_j \leq t} \left(\frac{n(t_j) - d(t_j)}{n(t_j)} \right). \tag{2}$$

The derivation of this equation may be found in Kalbfleisch and Prentice (2002) and also the original paper on the Kaplan–Meier estimator (Kaplan and Meier, 1958). For continuously and steadily cooled droplet freezing experiments T is directly proportional to t. As such the survivor function for droplet freezing data as a function of temperature is

$$\hat{S}(T) = \prod_{T_j \leq T} \left(\frac{N_U(T_j) - N_F(T_j)}{N_U(T_j)} \right). \tag{3}$$

where $N_U(T)$ is the number of unfrozen droplets at temperature T and $N_F(T)$ is the number of droplets that freeze at temperature T. Note that $f(T)$ is equal to $1 - \hat{S}(T)$. Equation 3 has the advantage that it is valid for 'censored' data, so can be used for datasets where it is not known whether droplets freeze or not. In principle, droplets can also be added part-way through an experiment. Best practice is to avoid these sorts of situation experimentally. Nevertheless, the ability to account for them may conceivably be useful.

2.2 Confidence intervals of the Kaplan–Meier Estimator

As discussed above, confidence intervals are useful for visualising the likely values which fraction frozen may take. Standard approaches for the calculation of the variance of $\hat{S}(t)$ may yield values outside the possible range of 0 to 1. As such the confidence intervals for the Kaplan–Meier estimator are usually calculated by applying the asymptotic normal distribution to a transformation of $\hat{S}(t)$ for which the range is unrestricted. The same approach may be adopted for calculation of confidence intervals for $\hat{S}(T)$ and therefore $f(T)$. Assuming that the probability of the probability of freezing is reasonably approximated

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

by N_F/N_U and the probability of freezing at each timestep is independent, the asymptotic variance of $\ln[-\ln \hat{S}(T)]$ may be calculated as

$$\hat{\sigma}^2(T) = \frac{\sum_{T_j} \frac{N_F(T_j)}{N_U(T_j)\left(N_U(T_j)-N_F(T_j)\right)}}{\left\{\sum_{T_j} \ln\left(\frac{N_U(T_j)-N_F(T_j)}{N_U(T_j)}\right)\right\}^2} \quad (4)$$

where the sums are calculated between T and the first value of T at which a freezing event occurs (Kalbfleisch and Prentice, 2002). In cases where there is no censorship, as there won't be for most ice nucleation measurements, equation 4 reduces to

$$\hat{\sigma}^2(T) = \frac{1-\hat{S}(T)}{N_0(\hat{S}(T)(\ln \hat{S}(T))^2)}. \quad (5)$$

The confidence intervals for $\hat{S}(T)$ are then given by $[\hat{S}(T)^{\exp(+z_{\alpha/2}\hat{\sigma}(T))}, \hat{S}(T)^{\exp(-z_{\alpha/2}\hat{\sigma}(T))}]$ where $z_{\alpha/2}$ is the inverse of the standard normal cumulative distribution, which takes values of 1.96 for a 95% confidence interval and 1.64 for a 90% confidence interval. The confidence interval for $f(T)$ may be found as $[1-\hat{S}(T)^{\exp(+z_{\alpha/2}\hat{\sigma}(T))}, 1-\hat{S}(T)^{\exp(-z_{\alpha/2}\hat{\sigma}(T))}]$. A difficulty with this approach is that no confidence interval can be calculated when $f(T) = 1$ because the log-log transform is undefined at $\hat{S}(T) = 0$. We suggest assuming this final interval is the same as the penultimate interval and have adopted this approach for the figures in this paper. We note that we calculated confidence intervals for $f(T)$ in the manner described here in Kinney et al. (2024).

Droplet fraction frozen data is often used to calculate surface density of sites active above temperature T, $n_s(T)$ to allow comparison of measurements conducted on different instruments or in different concentration regimes. This can be calculated as

$$n_s(T) = \frac{-\ln(1-f(T))}{\mathrm{A}} \quad (6)$$

Where A is the total surface area of INPs in a sample unit, usually given in cm^2 (Vali et al., 2015). Other quantities, for instance mass, can be used to normalise instead of A. To calculate the confidence interval for $n_s(T)$ equation 6 may be applied, replacing $f(T)$ with its confidence interval. It is worth noting that the abovementioned difficulty with the confidence interval of the final freezing droplet is avoided as $n_s(T)$ also cannot be calculated when $f(T) = 1$.

The approach to calculating confidence intervals is non-parametric in nature, meaning no assumption is made about the distribution of freezing temperatures. It shares this feature with the bootstrapping approach of Fahy et al. (2022) but is simpler to implement and understand, requiring no generation of synthetic experiments.

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

2.3 The log-rank test

A common question in the analysis of ice nucleation data is 'are these two fraction frozen curves different?'. For instance, an investigator may wish to establish whether a chemical treatment changes the nucleation temperature of a specific ice nucleator significantly or determine whether a measurement conducted in the presence of an ice nucleator is distinct from the background ice nucleation induced by the instrument. To date, such comparison has typically been conducted visually or by use of ad-hoc statistical comparison of curves. Here, we apply a standard approach for comparing survival curves (Kalbfleisch and Prentice, 2002) to the comparison of fraction frozen curves, constituting a more rigorous method for establishing whether the population of ice nucleators generating fraction frozen curves are different in nature. As with the Kaplan–Meier estimator the log-rank test breaks down the experiment into distinct events and looks at the set of droplets at risk immediately before a given event occurs. In the case of the Kaplan–Meier method, this set is used to calculate the conditional probability of a droplet surviving beyond temperature T. In contrast the log-rank test estimates the expected number of events at temperature T under the null hypothesis that two freezing curves are identical. It then compares the expected number of freezing events with the observed number to assess whether curves differ significantly.

Suppose that we want to test the null hypothesis that droplet freezing experiments represented by survivor functions $\hat{S}_0(T)$ and $\hat{S}_1(T)$ have identical underlying freezing probabilities. At each freezing temperature we first compute how many unfrozen droplets there are in each at temperature as

$$N_{U,tot}(T) = N_{U,o}(T) + N_{U,1}(T). \tag{7}$$

We then compute the total number of freezing events as

$$N_{F,tot}(T) = N_{F,o}(T) + N_{F,1}(T). \tag{8}$$

Under the null hypothesis, the expected number of freezing events in experiment 1 at temperature T is

$$E_1(T) = N_{U,1}(T)\frac{N_{F,tot}(T)}{N_{U_{tot}}(T)} \tag{9}$$

and the variance of the expected number of events at temperature T is

$$V_i(T) = \frac{N_{U,o}(T)N_{U,1}(T)\left(N_{U,tot}(T) - N_{F,tot}(T)\right)}{N_{U,tot}(T)^2\left(N_{U,tot}(T) - 1\right)}. \tag{10}$$

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

The test statistic for the log-rank test is then the sum of the squared difference between observed and expected events, normalised by the variance, summed across all steps (Kalbfleisch and Prentice, 2002).

$$\chi^2 = \frac{\left(\sum N_{F,1} - E_1\right)^2}{\sum V_i} \tag{11}$$

The χ^2 statistic with one degree of freedom is used to calculate the p-value (Kalbfleisch and Prentice, 2002). This may be done using statistical tables or (for example) the scipy function stats.chi2.sf, giving a formal measure of the probability that the differences between the two freezing curves occurred by chance. Because the numerator of the test statistic is the sum of differences you need enough droplets in the experiment to assume that central limit theorem applies.

While the log-rank test is non-parametric and hence makes no assumption about the distribution of freezing temperatures it is important to note that the log-rank test assumes proportional hazards, meaning that the ratio of freezing probabilities between two curves is constant across the entire temperature measurement range. In practice, this means that if experimental $f(T)$ curves cross the log-rank test loses statistical power because differences at warmer temperatures may effectively cancel out differences at colder temperatures.

3 Results and discussion

To demonstrate use of Kaplan–Meier statistics and the log-rank test we now present the results of data analysis using these methods both randomly generated simulated datasets, and some laboratory data from literature. The analysis was performed using a Jupyter notebook available on GitHub.

3.1 Comparison of KM confidence intervals with literature confidence intervals

Fahy et al. (2022) proposed a method for calculating confidence intervals on droplet freezing data using a variety of related methods, arguing that a 'studentized' confidence interval produced via empirical bootstrapping is the most accurate. In this approach the observed freezing temperatures are resampled and smoothly interpolated to create functions of $n_s(T)$. Confidence intervals are then taken from the bootstrapped ensemble. Figure 1 shows data reported in figure 4 of Fahy et al. (2022). Briefly, this droplet freezing data represents 91 100 nL droplets containing 0.1 wt% of an unaged volcanic ash first studied in Jahn et al. (2019). The experiment was conducted using the CMU-CS droplet-on-substrate system (Polen et al., 2018).

In Figure 1(a) we show $f(T)$ for this data with 95% KM confidence intervals and also 95% confidence intervals produced using a Monte-Carlo approach (called parameterised bootstrapping in Fahy et al. (2022)) based on the work of Vali (2019) and

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

first reported in the supplementary information of Whale (2022). Briefly, freezing events were binned by temperature, bin counts taken as expectation values and 5,000 Poisson draws per bin used to generate synthetic fraction-frozen curves. Confidence intervals represent the 2.5th–97.5th percentile envelope of the simulated fraction-frozen values. Figure. 1(b) shows KM, binned Monte-Carlo and studentized confidence intervals of $n_s(T)$ for the data. Encouragingly, all three types of confidence interval are broadly similar and seem reasonable. As discussed by Fahy et al. (2022) binning droplet freezing data is in some sense 'inefficient' as data points are combined. We note also that the method does not work well when there are relatively few droplets in a bin. This tends to be a particular issue when confidence intervals for $f(T)$ are transformed into $n_s(T)$, as can be seen in the first and last two points for the Monte-Carlo confidence intervals in Figure 1(b) where unreasonable lower and upper confidence intervals, respectively, are produced. Pragmatic use of the confidence intervals can easily mitigate this deficiency although it is a helpful that the KM and studentized confidence intervals avoid this difficulty.

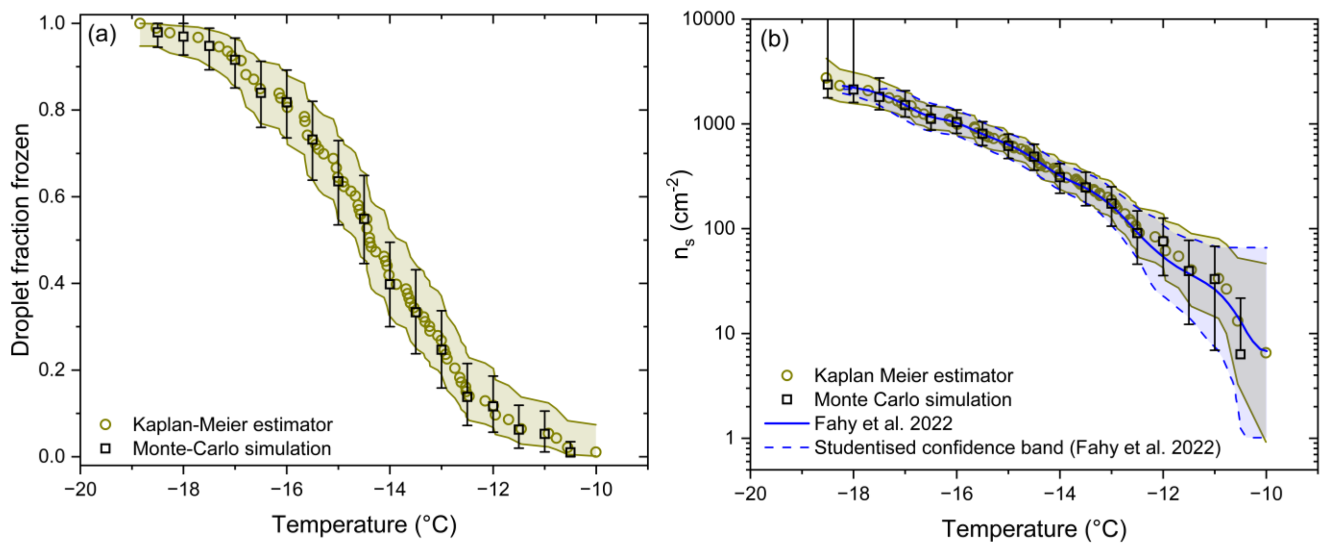

Figure 1. Comparison of Kaplan–Meier, Binned Monte-Carlo and Studentized 95% confidence intervals on (a) droplet $f(T)$ and (b) $n_s(T)$ for volcanic ash droplet freezing data from Fahy et al. (2022).

3.2 Simulated datasets for comparison of confidence intervals with varying numbers of droplets and slopes of fraction frozen

To generate datasets well suited to illustrating the properties of the Kaplan–Meier estimator and the log-rank test, we used a random-number generator to create plausible synthetic data resembling that typically obtained in droplet-freezing experiments. Twelve experiments were simulated using the Gaussian Random Number at https://www.random.org/gaussian-distributions/. In doing this, we assume a parametric form for the underlying probability density function (PDF) of freezing temperatures which is likely physically reasonable in at least some situations (De Almeida Ribeiro et al., 2023). Under this assumption,

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

different choices of the Gaussian parameters imply different conditional freezing probabilities as a function of temperature (i.e., different freezing "hazard" profiles). Because the log-rank test is sensitive to differences in these conditional freezing probabilities, it provides a reasonable approach for assessing whether two samples are consistent with the same underlying freezing behaviour. Table 1 reports the underlying means (μ) and standard deviations (σ) of the random samples generated.

Table 1. Summary statistics for simulated droplet freezing data produced using the Gaussian Random Number at random.org. Column headings give mean and standard deviation of the distribution from which the datasets were drawn the while the data in the table gives the actual means and standard deviation of the datasets.

Number of droplets	μ = -15°C σ = 1°C	μ= -15°C σ = 3°C
5	-15.50 (0.64)	-13.14 (2.48)
20	-14.54 (1.02)	-13.99 (3.14)
50	-15.128 (0.89)	-15.17 (3.20)
100	-15.068 (1.10)	-15.1342 (2.73)
1000	-14.9561 (1.03)	-14.861 (2.99)

Figure 2 shows comparison of simulated fraction frozen curves with underlying mean = -15°C and σ = 1°C dataset with various droplet numbers. Figure 2(a) compares simulated $f(T)$ curves for a very low number of droplets (5) against a more typical number (50) and a larger-than-typical number of droplets (1000). The confidence intervals with five droplets are broad enough to render the experiment meaningless. This is partially because the droplet number is too small for the assumptions required to derive equations 4 and 5 to be satisfied. The analysis therefore probes a parameter regime in which the method is not expected to be valid. An interesting feature here is that for our randomly generated 1000 droplet dataset the confidence interval sits above the cumulative normal distribution the sample is derived from. Clearly, this is not that unlikely an occurrence, but does highlight the care that should be taken, even with experiments that freeze relatively large numbers of droplets. Figure 2(b) shows simulated experiments with droplet numbers in more typical ranges for droplet freezing experiments. Figure 2(c) shows the confidence intervals for simulated 50 droplet $f(T)$ curves with μ = -15°C and σ = 1°C and 3°C. The 'flatter' curve produces a broader-looking confidence interval. The y-axis span of the confidence interval is identical for both curves (compare early and late points on the curves) but the steep curve compresses the intervals in the x-axis of the graph. Figure 2(d) shows a similar effect for typical numbers of droplets used in droplet freezing experiments, showing increased apparent width compared to the analogous steeper curves in Figure 2(b). Overall, the number of droplets required to constitute an acceptable experiment is essentially a matter of taste, however flatter $f(T)$ curves might generally require more droplets to be frozen compared to steeper $f(T)$ curves.

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

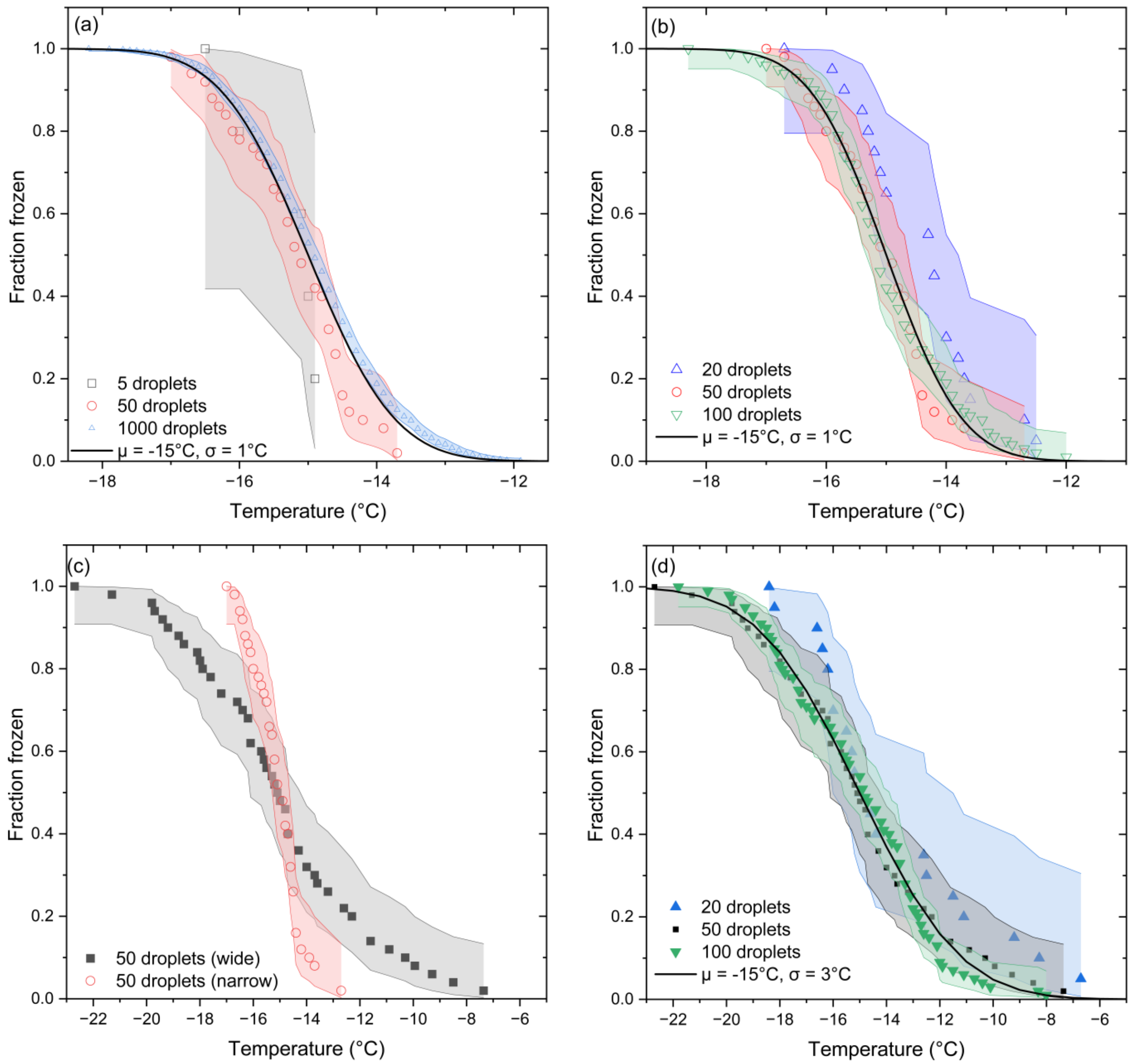

Figure 2. Comparison of 95% log-log Kaplan–Meier confidence intervals for randomly generated fraction frozen curves. (a) 240 compares experiments with very small and large droplets numbers with a typical experiment and the underlying normal distribution. (b) compares typical experimental droplet numbers, again with the underlying normal distribution. (c) compares steeper and flatter fraction frozen curves. (d) compares the typical experimental droplet numbers for the flatter fraction frozen curve.

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

3.3 Applications of the log-rank test

We now present a couple of examples of application of the log-rank test to ice nucleation data. Figure 3 shows $f(T)$ data for 1 µL droplets data from Daily et al. (2022) where a 0.5% w/v birch pollen washing water (BPWW) solution was subjected to heat treatments. Fresh, untreated BPWW nucleates ice with a very steep $f(T)$ curve. Table 2 reports log-rank tests comparing the measurement pairs shown in Figure 3. Wet-heated (i.e., boiled) BPWW nucleates at slightly cooler temperatures than untreated BPWW, a difference the log-rank test deems significant. In contrast, dry heating greatly reduces BPWW ice nucleation activity; its behaviour is not significantly different from the handling blank (a sample of purified water heat treated in the same way as the BPWW sample) as determined by log-rank test (Table 2). While the conclusions of Daily et al. (2022) were similar, with the important caveat of no claimed difference between the untreated and wet heated cases, applying the log-rank test provides a more quantitative assessment of the results' statistical significance.

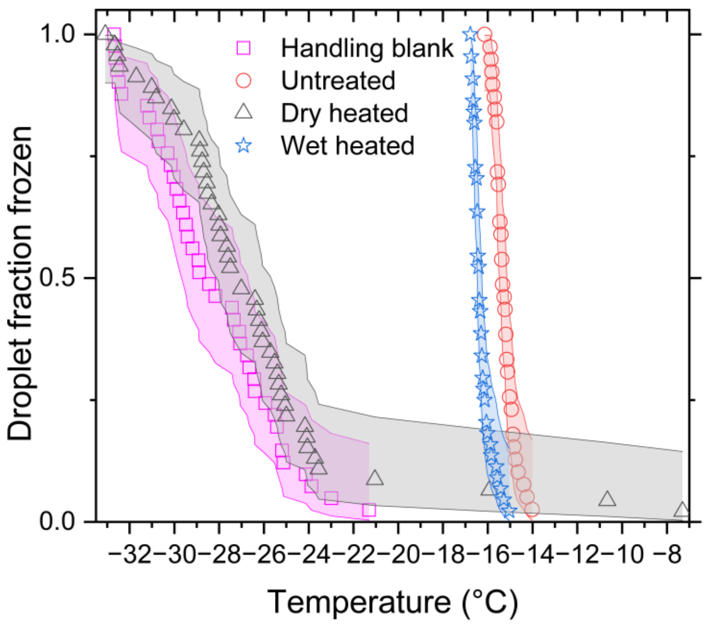

Figure 3. Fraction frozen curves for untreated and heat treated 0.5% w/v birch pollen washing water solution, and water handling blank for the dry heating procedure.

Table 2. Pairwise log-rank tests for $f(T)$ data for Birch pollen washing water solutions from Daily et al. (2022).

$f(T)$ Curve 1	$f(T)$ curve 2	Log-rank χ2	Log-rank p-value	Significant? (p<0.05)
Handling blank	Untreated BPWW	96.56	8.65×10^{-23}	Yes
Handling blank	Dry heated BPWW	1.78	0.18183	No
Handling blank	Wet heated BPWW	97.75	4.74×10^{-23}	Yes
Untreated BPWW	Dry heated BPWW	89.37	3.28×10^{-21}	Yes
Untreated BPWW	Wet heated BPWW	75.51	3.63×10^{-18}	Yes
Dry heated BPWW	Wet heated BPWW	85.12	2.80×10^{-20}	Yes

We now use some of the simulated the datasets reported in table 1 and a 100-droplet simulated dataset drawn from an underlying normal distribution with μ = -16°C and σ = 1°C to examine the sensitivity of the log-rank test under several scenarios and to highlight its limitations for ice-nucleation measurements. Figure 4(a) compares the 100 droplet simulations with underlying means of −15°C and −16°C and σ = 1°C. As expected, the log-rank test detects a clear difference. Figure 4(b) shows fraction-frozen curves drawn from the same distribution (μ = -15, σ = 1) that are nevertheless judged different, incorrectly rejecting the null hypothesis. This appears to be a Type I error due to sampling variability. As can be seen in table 1 the simulated 20 droplet curve has an improbable high mean (-14.54°C), leading to this error. In Figure 4(c), the null is not rejected even though the underlying populations differ; here the curves cross, violating the proportional-hazards assumption on which the log-rank test relies and reducing power. Figure 4(d) provides a counterexample in which, despite a crossing, the null is correctly rejected. For most ice nucleation studies, such crossings are unlikely, but the possibility should be considered. In the medical statistics literature, methods for handling crossing survival curves have been explored (Dormuth et al., 2022) and could be adapted to droplet-freezing data if needed.

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

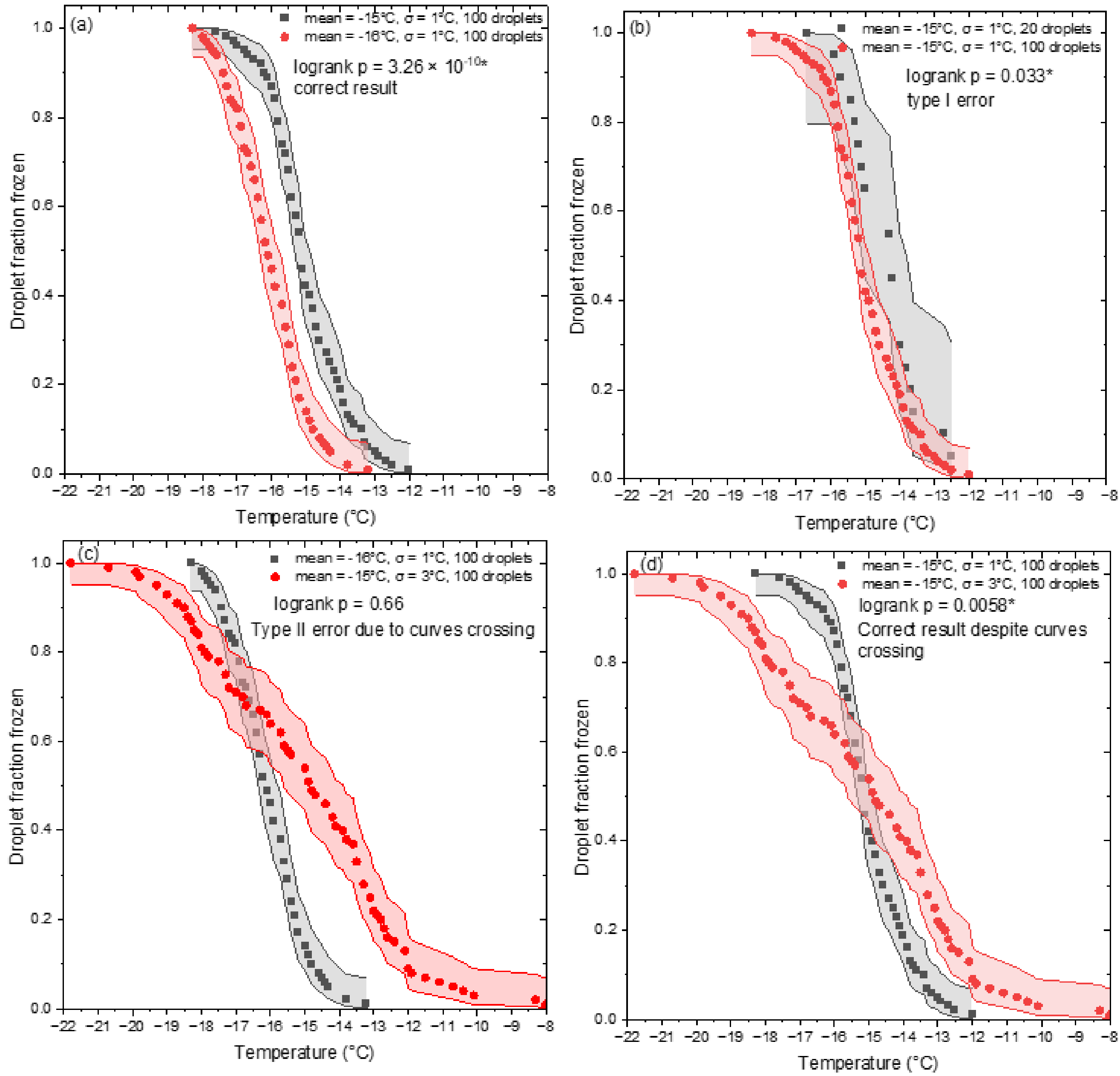

Figure 4. Simulated fraction frozen curves evaluated with the log-rank test. Asterisks (*) mark significant results at $\alpha = 0.05$ (a) Curves with means of −15°C vs −16°C; the test correctly detects a difference. (b) Identical underlying distribution but different droplet numbers, yielding a Type I error (false positive). (c) Curves with different slopes that cross, yielding a Type II error (false negative). (d) Overlapping, crossing curves that nonetheless differ overall; the test returns a correct significant result.

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

4 Conclusions

Framing droplet-freezing measurements as survival data enables direct use of the Kaplan–Meier (KM) estimator to provide a simple, reproducible method for calculating nonparametric confidence intervals (CIs) for droplet fraction frozen ($f(T)$) and derived $n_s(T)$ values. On a literature volcanic-ash dataset, KM CIs broadly agree with Monte-Carlo and studentized-bootstrapped CIs while avoiding binning-related edge artifacts, supporting KM as a robust option for uncertainty quantification. KM confidence intervals show that precision depends on sample size: very small experiments yield overly wide CIs, whereas ~50 droplets or more typically provide usable constraints for $f(T)$ and $n_s(T)$, depending somewhat on the slope of $f(T)$. Compared to previously published methods for calculating CIs for droplet freezing data our approach has the advantages of being analytical in nature and relatively simple to understand and implement. We recommend reporting $f(T)$ and $n_s(T)$ with KM CIs and the number of droplets analysed.

The log-rank test gives a rigorous hypothesis test for differences between fraction-frozen curves which we have applied to literature birch pollen washing water (BPWW) data. Application to simulated data highlights limitations. Sampling variability can of course yield errors, and crossings in $f(T)$ curves can mask true differences because the proportional-hazards assumption reduces power in these cases. Where $f(T)$ curves are compared we recommend calculating and reporting and log-rank statistics and p-values to improve transparency and comparability across studies.

Code and data availability. A current version of the analysis Jupyter notebook and input data used in this paper may be found at https://github.com/TFWhale/KM-nucleation-stats

Author contributions. TFW and SLB conceptualised the paper. TFW wrote the initial version code of the analysis Python script and performed the analysis presented in the paper. TSS checked and refined the statistical approaches employed and improved the Python script used. TFW wrote the initial draft. TFW, SLB and TSS edited the final version of the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements: An AI-based assistant (Microsoft Copilot) was used to support language editing and to provide general programming assistance during development of the analysis notebook. The analytical approach, implementation choices, and validation were performed by TFW.

Funding sources. TFW was supported by Natural Environment Research Council grant 'From Toy to Cloud Modelling: Leveraging Molecular Simulations to Improve Atmospheric Models of Ice Nucleation', reference NE/Z504245/1.

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

References

Agresti, A. and Coull, B. A.: Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions, The American Statistician, 52, 119-126, 10.1080/00031305.1998.10480550, 1998.

Bigg, E. K.: The supercooling of water, Proceedings of the Physical Society. Section B, 66, 688, 1953.

Daily, M. I., Tarn, M. D., Whale, T. F., and Murray, B. J.: An evaluation of the heat test for the ice-nucleating ability of minerals and biological material, Atmos. Meas. Tech., 15, 2635-2665, 10.5194/amt-15-2635-2022, 2022.

de Almeida Ribeiro, I., Meister, K., and Molinero, V.: HUB: a method to model and extract the distribution of ice nucleation temperatures from drop-freezing experiments, Atmos. Chem. Phys., 23, 5623-5639, 10.5194/acp-23-5623-2023, 2023.

Dormuth, I., Liu, T., Xu, J., Yu, M., Pauly, M., and Ditzhaus, M.: Which test for crossing survival curves? A user's guideline, BMC Med Res Methodol, 22, 34, 10.1186/s12874-022-01520-0, 2022.

Fahy, W. D., Shalizi, C. R., and Sullivan, R. C.: A universally applicable method of calculating confidence bands for ice nucleation spectra derived from droplet freezing experiments, Atmos. Meas. Tech., 15, 6819-6836, 10.5194/amt-15-6819-2022, 2022.

Harrison, A. D., Whale, T. F., Carpenter, M. A., Holden, M. A., Neve, L., O'Sullivan, D., Vergara Temprado, J., and Murray, B. J.: Not all feldspar is equal: a survey of ice nucleating properties across the feldspar group of minerals, Atmos. Chem. Phys. Discuss., 2016, 1-26, 10.5194/acp-2016-136, 2016.

Herbert, R. J., Sanchez-Marroquin, A., Grosvenor, D. P., Pringle, K. J., Arnold, S. R., Murray, B. J., and Carslaw, K. S.: Gaps in our understanding of ice-nucleating particle sources exposed by global simulation of the UK Earth System Model, Atmos. Chem. Phys., 25, 291-325, 10.5194/acp-25-291-2025, 2025.

Jahn, L. G., Fahy, W. D., Williams, D. B., and Sullivan, R. C.: Role of Feldspar and Pyroxene Minerals in the Ice Nucleating Ability of Three Volcanic Ashes, ACS Earth and Space Chemistry, 3, 626-636, 10.1021/acsearthspacechem.9b00004, 2019.

Kalbfleisch, J. D. and Prentice, R. L.: The statistical analysis of failure time data, Second edition., Wiley series in probablility and statistics, Wiley, New York ;2002.

Kaplan, E. L. and Meier, P.: Nonparametric Estimation from Incomplete Observations, Journal of the American Statistical Association, 53, 457-481, 10.1080/01621459.1958.10501452, 1958.

Kinney, N. L. H., Hepburn, C. A., Gibson, M. I., Ballesteros, D., and Whale, T. F.: High interspecific variability in ice nucleation activity suggests pollen ice nucleators are incidental, Biogeosciences, 21, 3201-3214, 10.5194/bg-21-3201-2024, 2024.

Knopf, D. A. and Alpert, P. A.: Atmospheric ice nucleation, Nature Reviews Physics, 5, 203-217, 10.1038/s42254-023-00570-7, 2023.

Koop, T., Luo, B. P., Biermann, U. M., Crutzen, P. J., and Peter, T.: Freezing of HNO3/H2SO4/H2O solutions at stratospheric temperatures: Nucleation statistics and experiments, Journal of Physical Chemistry, 101, 1117-1133, 10.1021/jp9626531, 1997.

Miller, A. J., Brennan, K. P., Mignani, C., Wieder, J., David, R. O., and Borduas-Dedekind, N.: Development of the drop Freezing Ice Nuclei Counter (FINC), intercomparison of droplet freezing techniques, and use of soluble lignin as an atmospheric ice nucleation standard, Atmos. Meas. Tech., 14, 3131-3151, 10.5194/amt-14-3131-2021, 2021.

Murray, B. J., O'Sullivan, D., Atkinson, J. D., and Webb, M. E.: Ice nucleation by particles immersed in supercooled cloud droplets, Chemical Society Reviews, 41, 6519-6554, 10.1039/C2CS35200A, 2012.

Polen, M., Brubaker, T., Somers, J., and Sullivan, R. C.: Cleaning up our water: reducing interferences from non-homogeneous freezing of "pure" water in droplet freezing assays of ice nucleating particles, Atmos. Meas. Tech. Discuss., 2018, 1-31, 10.5194/amt-2018-134, 2018.

Vali, G.: Quantitative Evaluation of Experimental Results an the Heterogeneous Freezing Nucleation of Supercooled Liquids, Journal of the Atmospheric Sciences, 28, 402-409, 10.1175/1520-0469(1971)028<0402:qeoera>2.0.co;2, 1971.

Vali, G.: Revisiting the differential freezing nucleus spectra derived from drop-freezing experiments: methods of calculation, applications, and confidence limits, Atmos. Meas. Tech., 12, 1219-1231, 10.5194/amt-12-1219-2019, 2019.

Vali, G., DeMott, P. J., Möhler, O., and Whale, T. F.: Technical Note: A proposal for ice nucleation terminology, Atmos. Chem. Phys., 15, 10263-10270, 10.5194/acp-15-10263-2015, 2015.

Whale, T. F.: Disordering effect of the ammonium cation accounts for anomalous enhancement of heterogeneous ice nucleation, The Journal of Chemical Physics, 156, 144503, 10.1063/5.0084635, 2022.

https://doi.org/10.5194/egusphere-2026-680
Preprint. Discussion started: 9 March 2026

Whale, T. F., Murray, B. J., O'Sullivan, D., Wilson, T. W., Umo, N. S., Baustian, K. J., Atkinson, J. D., Workneh, D. A., and Morris, G. J.: A technique for quantifying heterogeneous ice nucleation in microlitre supercooled water droplets, Atmos. Meas. Tech., 8, 2437-2447, 10.5194/amt-8-2437-2015, 2015.