Supplement to:

Reconciling Arctic sea ice roughness across four observing platforms: a resolution transfer function and the stability of spectral shape over six decades

Lei Yang and Lin Zhu

This supplement contains the robustness analyses supporting the main text. The values reported here are those produced by the analysis pipeline described in the Methods section of the main text. Sections S1–S2 establish that the broken power law is a real feature of the spectra rather than an artefact of the fitting procedure. Section S3 shows that the resolution transfer function follows from the spectral model. Sections S4–S7 quantify the robustness of the trend and Hurst-exponent results against clustering, instrument resolution, region, and segment length. Sections S8–S9 address melt-season contamination and the stability of the ICESat-2 record.

S1 The break is not an artefact of fitting

A single power law fitted with the same broken-power-law routine will always report some slope change, because the routine is free to place a break anywhere. To calibrate that, we generated matched single-power-law surrogate profiles ($\beta = -1.69$) with the same length and sampling as the observations and passed them through the identical pipeline.

The observed slope change at the break is 2.31 times the null value (Fig. S1a), and the two distributions are separated far beyond chance (Mann–Whitney $p < 10^{-100}$). The break location discriminates even more sharply (Fig. S1b): observed breaks cluster at $\lambda_{\mathrm{break}} = 79$ m with 64 % falling in the physically meaningful 50–200 m ridge-spacing band, whereas the null breaks collapse onto the analysis-band edge at 256 m with only 10 % in that band. This comparison uses the 989-segment subset for which matched surrogates were generated, so its median break differs from the 99 m reported in the main text for the full 43-cruise set; what matters here is the contrast between observed and null breaks, not the absolute value. A surrogate spectrum with no break is thus pushed to the edge of the fitting window, while the real spectra place the break at a consistent interior scale.

S2 Sensitivity to the number of spectral bins

The spectra are log-binned before the break search, so the bin count is a free analysis choice. Repeating the null comparison for every bin count from 20 to 34 leaves the observed-to-null slope-change ratio between 2.21 and 2.25 (Fig. S2), a spread of under 2 % and far from the value of 1 that would indicate no detectable break. The 21 bins used in the main text are not a special case.

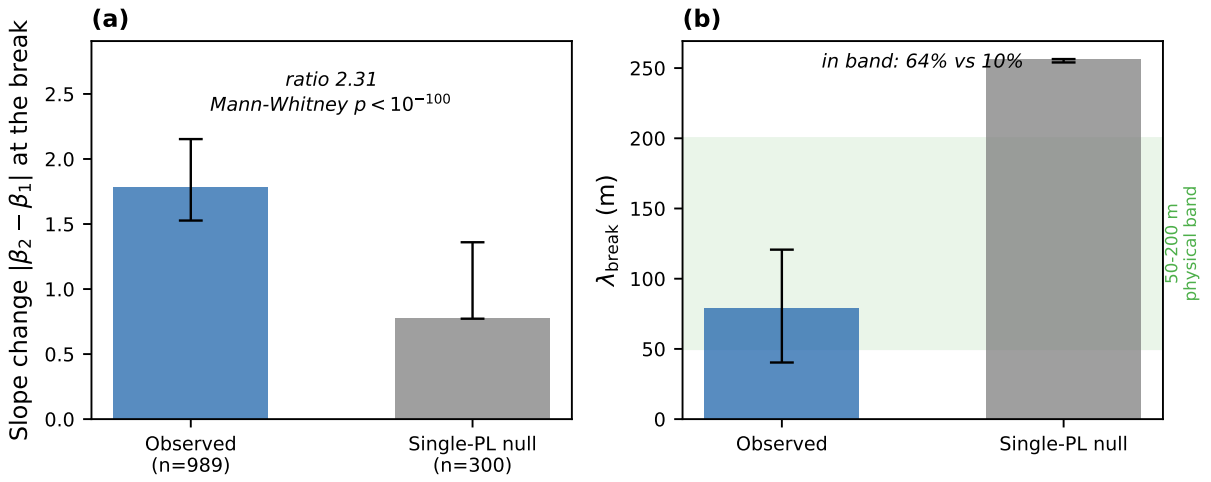


Figure S1: Broken power law versus a matched single-power-law null. (a) Median slope change $|\beta_2 - \beta_1|$ recovered by the fitting routine for the observed profiles ($n = 989$) and for surrogate single-power-law profiles ($n = 300$). Error bars show the interquartile range for the observations and the median-to-95th-percentile range for the null. (b) Recovered break wavelength. Shading marks the 50–200 m band expected for ridge spacing; the null breaks pile up at the analysis-band edge.

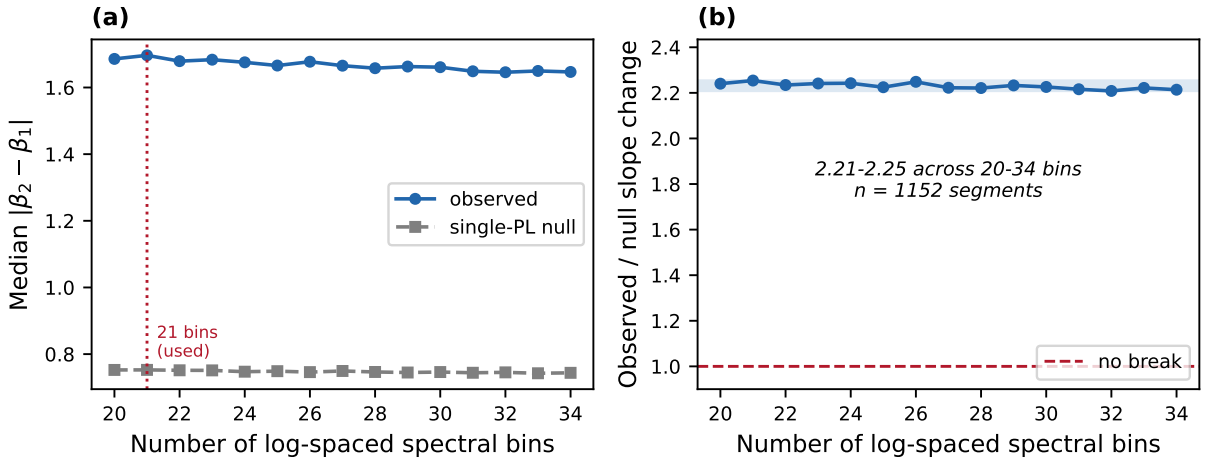


Figure S2: Bin-count sensitivity of the break diagnostic, computed on 1152 submarine segments. (a) Median slope change for the observations and the matched null. (b) Their ratio, with the range across all tested bin counts shaded. The dashed line marks the no-break expectation.

S3 The degradation curve follows from the spectral model

The transfer function reported in the main text ($z_2/z_1 = 0.218\,\Delta x^{-0.46}$) is an empirical fit to the smoothing experiment, but its exponent is predicted by the spectral model. Numerically integrating Eq. (3) of the main text over the observed broken power law ($\beta_1 = -0.505$, $\beta_2 = -2.087$, $\lambda_{\mathrm{break}} = 98.6$ m) gives $z_2/z_1 = 0.200\,\Delta x^{-0.484}$ with $R^2 = 0.9995$, and the high-resolution asymptotic limit $-(\beta_2+3)/2$ gives -0.457 (Fig. S3). The empirical exponent lies between the two. The degradation curve is therefore a consequence of the spectral shape rather than an independent empirical relation, which is why it can be applied to platforms outside the calibration set.

The absolute offset between the predicted and observed curves reflects the idealised smoothing kernel used in the integral; the exponent, which is what the transfer function relies on, is insensitive to that choice.

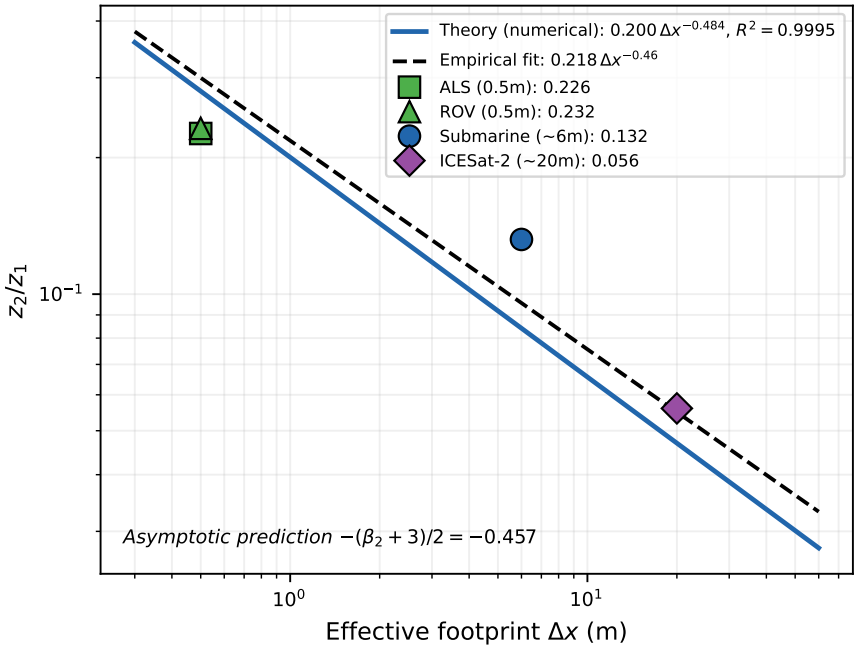


Figure S3: Resolution degradation curve predicted from the broken power law (solid) compared with the empirical fit used in the main text (dashed) and the four observed platform medians. Predicted values use a Gaussian sensor kernel.

S4 Hurst exponent under cluster resampling

Segments within one submarine cruise are strongly autocorrelated, so treating them as independent would understate the uncertainty on the Hurst exponent. Resampling whole cruises (10 000 bootstrap replicates over 43 cruises, 5470 segments) gives $H = 0.59$ with a 95 % interval of 0.53–0.73. That interval clears both a Brownian profile ($H = 0.5$) and the low-exponent brittle-fracture class reported for compacted-grain sandstone ($\zeta = 0.46 \pm 0.05$, i.e. 0.41–0.51) (Fig. S4). Weighting every cruise equally instead gives $H = 0.54$ with an interval of 0.50–0.63, which touches the Brownian value and overlaps the upper end of the sandstone interval. Neither exclusion is therefore weighting-independent: sea ice sits above both reference classes under either scheme, but only the cluster-bootstrap estimate separates it from them with confidence. The main text states the comparison with that restriction.

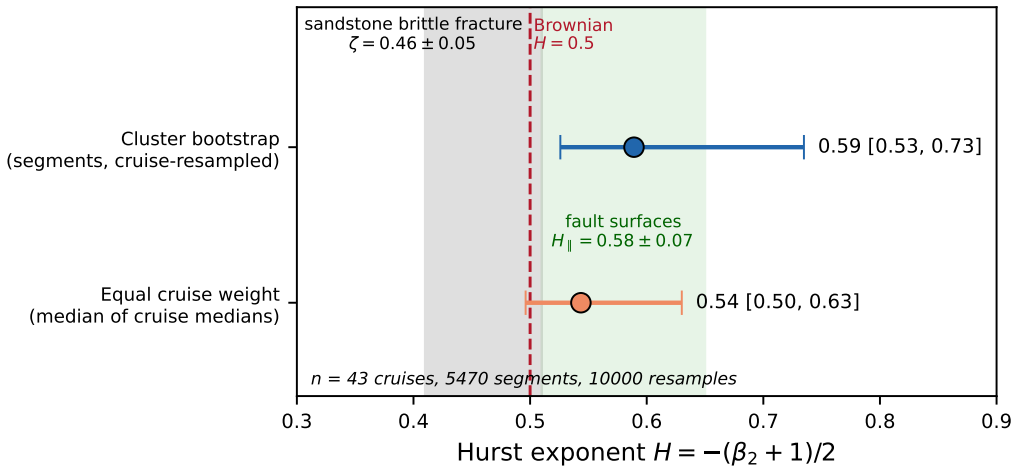


Figure S4: Hurst exponent of the high-frequency spectral regime under two weighting schemes, with 95 % confidence intervals. Reference bands mark the low-exponent brittle-fracture class of compacted-grain sandstone ($\zeta = 0.46 \pm 0.05$), the Brownian value, and the slip-parallel roughness of fault surfaces ($H_{\parallel} = 0.58 \pm 0.07$).

S5 Instrument resolution does not generate the temporal result

Native sampling interval d_x differs among cruises and correlates with all three spectral parameters, most strongly with λ_{break} ($r = +0.76$). If d_x had drifted over the record it could manufacture or mask a trend. It did not ($r = -0.14$, $p = 0.36$ against year), and partial correlations controlling for d_x leave the conclusions unchanged (Fig. S5): β_1 and β_2 remain non-significant, while λ_{break} retains the marginal signal discussed in the main text, which then fails once the two 1960 leverage cruises are removed.

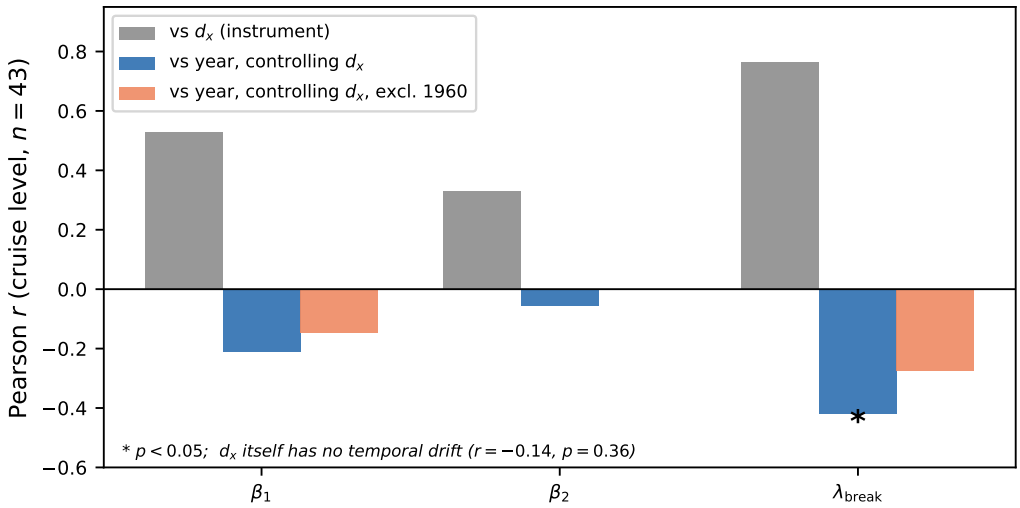


Figure S5: Cruise-level correlations ($n = 43$). Grey: each spectral parameter against the native sampling interval d_x, showing the instrumental dependence. Blue and orange: correlation with year after partialling out d_x, for the full sample and excluding the two 1960 cruises.

S6 Regional sampling

Early cruises favoured different Arctic sectors than later SCICEX missions, so region is confounded with time. Grouping segments into seven region–ice-type classes gives a marginally significant difference in spectral deviation (ANOVA $p = 0.008$), but this is carried entirely by the two smallest and least well characterised groups (Chukchi first-year ice, $n = 46$, of which 31 segments lack a season assignment, and the residual OTHER class). Dropping those two leaves no detectable regional difference ($p = 0.91$), and among multi-year ice regions alone there is none either ($p = 0.67$; Fig. S6).

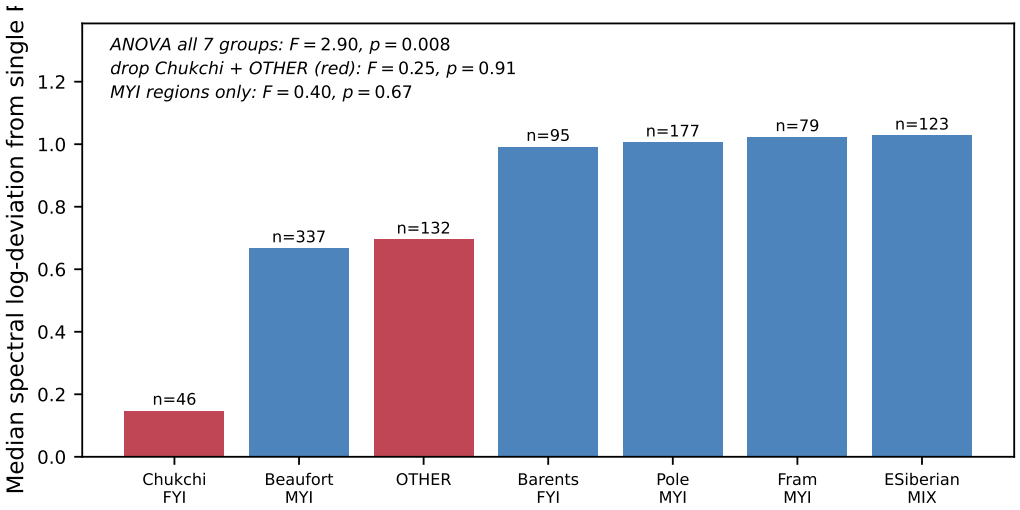


Figure S6: Median spectral deviation from a single power law by region and ice type. Red bars mark the two groups that carry the marginal overall ANOVA result. Sample sizes are shown above each bar.

S7 Segment length is a bandwidth effect

Longer segments resolve lower frequencies and therefore yield systematically smaller z_2/z_1, from 0.137 in the shortest length quartile to 0.099 in the longest (Fig. S7). This is the same bandwidth mechanism as the sensor resolution effect, acting on the lower integration limit rather than the upper one, and it is a further reason to base the temporal and cross-platform comparisons on the resolution-independent exponents.

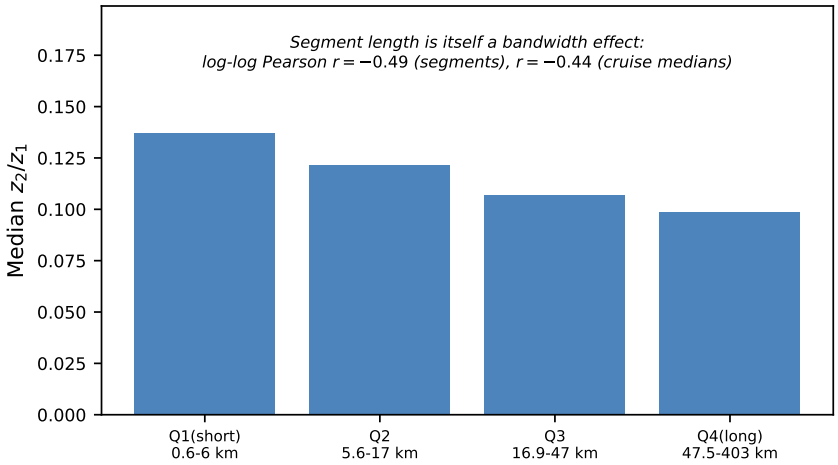


Figure S7: Median z_2/z_1 by segment-length quartile, with the length range of each quartile shown. Correlation statistics are given at segment level and at cruise-median level.

S8 Melt-pond contamination of the ALS surface

Melt ponds present a flat, specular surface that biases the ALS-derived freeboard. Progressively filtering low-freeboard pixels from the melt-season flights lowers the median z_2/z_1 monotonically (Fig. S8), but the melt-season elevation of z_2/z_1 survives even the most aggressive filter ($h_{\mathrm{fb}} > 0.15$ m), which is the threshold used in the main text. Pond contamination therefore accounts for part but not all of the melt-season signal, consistent with the independent evidence from the paired ROV comparison that the surface genuinely sharpens relative to the underside in summer.

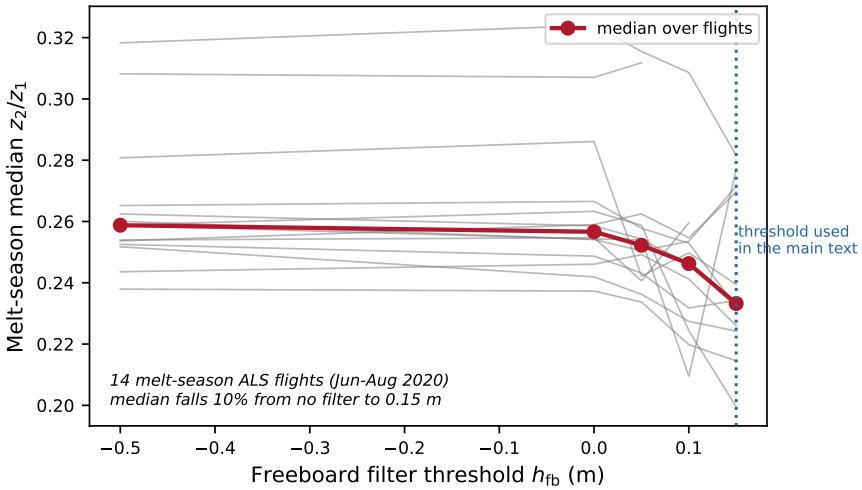


Figure S8: Melt-season ALS z_2/z_1 as a function of the freeboard filter threshold. Grey lines are individual flights; the red line is the median across flights. The dotted line marks the threshold adopted in the main text.

S9 Stability of the ICESat-2 record

Seven freeze-season sampling windows spanning five seasons (2018–2023, 8878 transects) show z_2/z_1 confined to 0.057–0.077 while the roughness amplitude z_1 varies more than eightfold between windows (Fig. S9). The decoupling of shape from amplitude documented in the submarine record therefore also holds in the satellite era, although the ICESat-2 record is far too short to constitute an independent trend test and is not used as one in the main text.

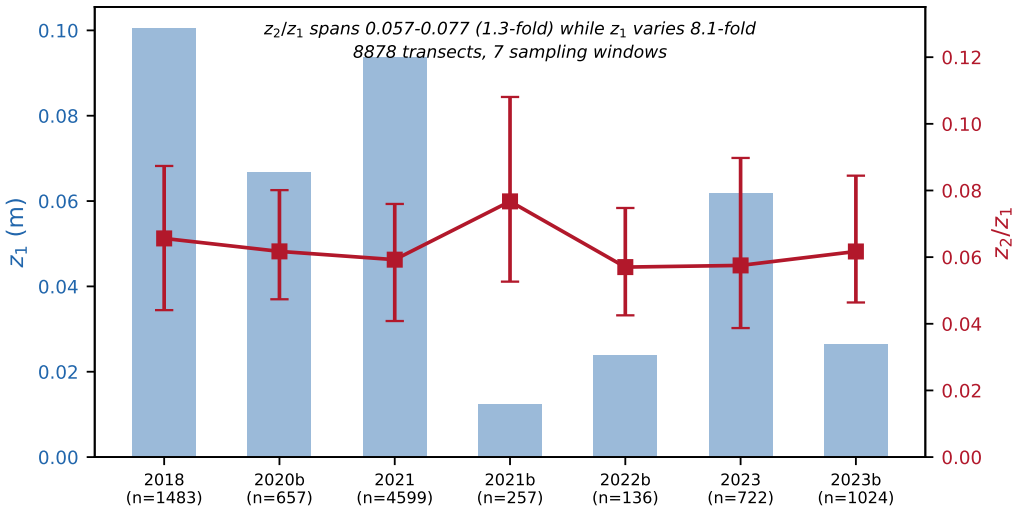


Figure S9: ICESat-2 ATL07 roughness amplitude z_1 (bars, left axis) and morphology ratio z_2/z_1 (line with interquartile range, right axis) for seven freeze-season windows. Labels ending in "b" denote the October–November window of that year; the remainder are January–February windows.