Review #1

Summary

The paper builds a continuous 20-year (2003-2023) record of coastal sea ice phenology at the TUKT ground station on the Beaufort Sea coast using GNSS Interferometric Reflectometry (GNSS-IR). The central methodological device is the Amplitude Integration Factor (AIF), which integrates reflected signal power across the 5–15° elevation range as a proxy for ice presence, together with a normalization scheme intended to bridge a 2015 receiver upgrade. Phenological transition dates (ice-out, freeze-up) are extracted from the calibrated AIF and used to derive interannual trends. The primary results are a shortening of the ice season of ~4.6 days per decade, an asymmetric decline dominated by delayed autumn freeze up rather than advanced spring breakup, strong coupling of freeze-up to accumulated Freezing Degree-Days (R^2 = 0.74), and a systematic ~5.5-day lead of GNSS-IR over the 4-km IMS product at freeze-up, interpreted as detection of nearshore frazil ice. This review targets the use of the data, the calibration and detection methodology, the statistical inference, the cross-scale comparisons, and the conclusions drawn. I do not focus on the internal details of the GNSS-IR signal processing chain (wavelet power extraction, Lomb-Scargle screening), which lie outside my primary expertise. The topic is a good fit for The Cryosphere and the dataset a continuous, high-completeness, sub-kilometre ground record over two decades. The physical reasoning is generally coherent and the multi-source 2019 case study is a nice demonstration.

Reservations stem from the verbosity of the written claims, and how strongly they are stated relative to the evidence. Several of the paper's central claims rest on a single significant p-value obtained after excluding one year, correlation windows selected to maximise fit, and detection rules that are asymmetric. Additionally, much more care should be taken to check for grammatical errors and consistency of reasoning, formatting and style, as it would benefit from a thorough editing pass for grammar, consistency and repetition before resubmission. The key questions to confront:

General response

We thank the reviewer for the careful and constructive assessment. We have completed the additional analyses described below and, if the editor invites a revised manuscript, will incorporate the corresponding textual, methodological, figure, and table changes. These planned revisions address the concerns regarding statistical framing, detector symmetry, receiver consistency, physical interpretation, cross-scale comparison, and presentation.

Most importantly, our reanalysis retains all valid years, including 2018. The primary phenological tendencies are estimated using Theil–Sen slopes and two-sided Mann–Kendall tests, with the period, sample size, 95 % confidence interval, and p value reported for each metric. None reaches the 0.05 significance level. In a revised manuscript, we will therefore describe these results as tendencies rather than statistically established changes and will use only the corrected cross-year ice-season duration.

We have also completed robustness analyses using 18 prespecified persistence and stability-confirmation configurations, a ±10 % state-threshold perturbation, four combinations of early and stable boundaries, and a formal matched-year comparison of ice-out and freeze-up trend magnitudes. Additional analyses include receiver-period diagnostics, empirical evaluation of the elevation-dependent AIF response, comparison with conventional peak amplitude, and window-search-adjusted significance tests for the degree-day analysis.

The IMS comparison was recalculated using four fixed neighbouring 4 km cells and is interpreted as a cross-scale comparison rather than independent validation. Interpretations concerning frazil ice, sensor resolution, and regional representativeness have been moderated. If invited to submit a revised manuscript, we will also reduce repetition and descriptive intensifiers, correct notation and formatting, shorten figure captions, and improve consistency throughout. Detailed responses and the locations of the planned manuscript changes are provided below.

Major Comments

Major comment 1: Statistical framing of the trends

Status: Addressed

Reviewer comment. *The abstract and conclusions lead with a "statistically significant shortening … 4.63 days per decade (p = 0.04)", a value that applies only to one metric (corrected ice duration) and only with 2018 excluded; with 2018 retained the same metric gives p = 0.06, and every other trend in Table 4 is non significant at 95% (ice-out p = 0.35, freeze-up p = 0.18, ice duration p = 0.21). Leading with the one number that crosses the threshold, without foregrounding that it required outlier removal, overstates the robustness. Please report all trends with confidence intervals up front, and consider a non-parametric test (Mann–Kendall / Theil Sen) that is better suited to n ≈ 19 with an extreme year and less sensitive to 2018.*

Response. We agree that the original manuscript placed too much emphasis on the corrected ice-duration trend obtained after excluding 2018. The reported shortening of 4.63 d decade^{-1} ($p=0.04$) depended on removing one year after examining the data and did not adequately represent the uncertainty in the complete record.

We therefore revised both the statistical analysis and the framing of the results. All years with valid observations, including 2018, are will retained in the primary analyses; no year is excluded because of its influence on the estimated trend or statistical significance. Theil–Sen slopes and two-sided Mann–Kendall tests, with correction for tied values, are used as the primary trend methods. Ordinary least-squares regression is no longer used as the primary method for the phenological trends and is retained only for explicitly identified complementary analyses.

Under the primary stable–stable boundary definition, stable ice-out tended to occur earlier at −1.97 d decade^{-1} (2004–2023, $n=20$; 95 % CI: −5.39 to +2.61 d decade^{-1}; $p=0.347$), whereas stable freeze-up tended to occur later at +2.85 d decade^{-1} (2003–2022, $n=20$; 95 % CI: −2.40 to +7.45 d decade^{-1}; $p=0.284$). The ice-free period tended to lengthen at +5.30 d decade^{-1} (2004–2022, $n=19$; 95 % CI: −4.54 to +12.06 d decade^{-1}; $p=0.294$), and corrected cross-year ice-season duration tended to shorten at −3.31 d decade^{-1} (2004–2023, $n=20$; 95 % CI: −7.82 to +0.63 d decade^{-1}; $p=0.086$). None of these primary trends reached the 0.05 significance level, and all confidence intervals included zero.

We also separated the primary results from the boundary-definition sensitivity analysis. Across the four combinations of early and stable transition boundaries, ice-free-period slopes ranged from +4.09 to +5.76 d decade^{-1}, and corrected cross-year ice-season-duration slopes ranged from −4.23 to −3.21 d decade^{-1}. The stable-ice-out/early-freeze-up combination produced a borderline corrected-duration estimate of −4.23 d decade^{-1} (95 % CI: −8.59 to −0.02 d decade^{-1}; $p=0.048$), but this result was not reproduced by the other three boundary definitions. It is therefore reported only as a sensitivity result and is not used as evidence of a statistically significant primary trend. The formal comparison of the ice-out and freeze-up trend magnitudes is reported in our response to Major Comment 2.

The former within-calendar-year ice-duration metric will be removed from all analyses, tables, figures and conclusions. All reported ice-season-duration results will use the corrected cross-year ice-season duration, defined as the continuous interval from freeze-up in year (y-1) to ice-out in year (y).

If invited to submit a revised manuscript, we will change the title from:

> **"Two decades of GNSS-IR observations reveal an asymmetric decline in coastal sea ice phenology of the Beaufort Sea"**

to:

> **"Two decades of GNSS-IR observations of coastal sea-ice phenology indicate contrasting ice-out and freeze-up tendencies at Tuktoyaktuk, Beaufort Sea."**

The revised title retains the observed tendencies toward earlier ice-out and later freeze-up while avoiding the implication that a statistically significant asymmetric decline has been established.

The revised Figure 9, reproduced below, now presents the complete-record transition-date and duration tendencies together with the sensitivity of the duration slopes to the four early/stable boundary combinations.

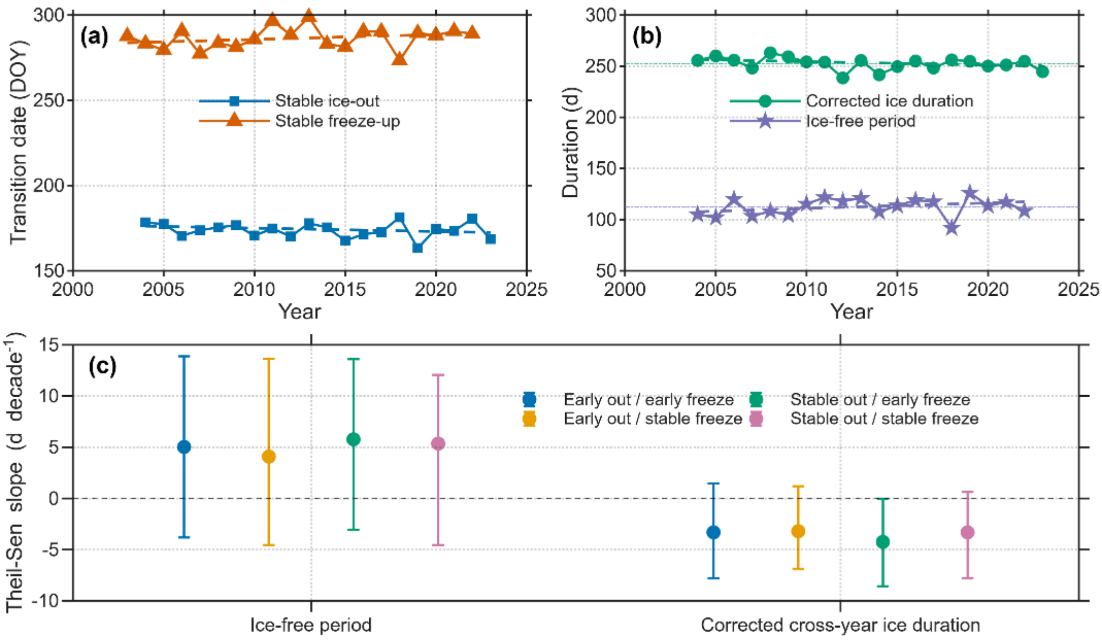


Figure 9. Long-term tendencies in GNSS-IR coastal sea-ice phenology at TUKT under the early- and stable-boundary framework. (a) Primary stable ice-out and stable freeze-up dates with their Theil–Sen trend lines. (b) Ice-free period and corrected cross-year ice-season duration under the primary stable–stable definition. Colored lines and markers show annual values, black dashed lines show Theil–Sen trends, and gray dash-dotted lines show record means. (c) Sensitivity of the duration slopes to the four combinations of early and stable transition boundaries. Error bars show two-sided 95 % confidence intervals, and the horizontal zero line indicates no temporal trend. Numerical estimates, sample sizes, confidence intervals and Mann–Kendall results are reported in Table 4.

Changes to be incorporated in the revised manuscript: The Title, Abstract, Introduction, trend-analysis methods in Sect. 3.3.3, Results and discussion in Sect. 4.2.2, Table 4, Figure 9 and Conclusions will be revised. All primary analyses will retain 2018, use Theil–Sen slopes and two-sided Mann–Kendall tests, and report metric-specific periods, sample sizes and 95 % confidence intervals.

Major comment 2: Asymmetric detection criteria

Status: Addressed

Reviewer comment. *It should be investigated as to whether the asymmetry observed in the freeze up / ice-out is a product of the detection criteria. The freeze-up rule requires the AIF to stay above 3σ for at least 5 days, whereas ice-out is the latest date the AIF drops below 3σ. Asymmetric persistence requirements can themselves bias freeze-up and ice-out in opposite directions and could induce part of the very asymmetry that is the paper's central finding. A sensitivity analysis on the detection/persistence rules additionally to the ±10% threshold test already reported, plus a formal test that the freeze-up and ice-out trend magnitudes are different, would help establish that the asymmetry is physical rather than methodological.*

Response. We agree that the detection rules in the original submission were asymmetric and could have affected the relative timing of ice-out and freeze-up. We therefore replaced them with equivalent, direction-reversed criteria. Under the revised primary detector, stable ice-out and freeze-up use the same state threshold, temporal-coverage requirement, preceding- and following-state support, longer-term confirmation, terminal seasonal anchor, sustained-reversal test, gap-handling rule, and candidate-de-duplication rule. Only the source and target states and their physically appropriate seasonal search intervals differ.

For both stable transitions, the primary detector requires at least 60 % coverage during the preceding and following 7 d periods, at least 50 % support for the preceding seasonal state, and at least 90 % support for the new state after the transition. Each candidate must also be supported by 14 equivalent observed days of the new state accumulated within no more than 35 calendar days and must pass the same terminal-anchor and 5 d sustained-reversal tests. Early ice-out and freeze-up are likewise detected symmetrically using a 3 d preceding-state window, a common 2 d persistence requirement, a preceding-state fraction of 0.70, an immediate target-state fraction of 0.90, and a minimum temporal coverage of 0.80. No transition date is assigned manually.

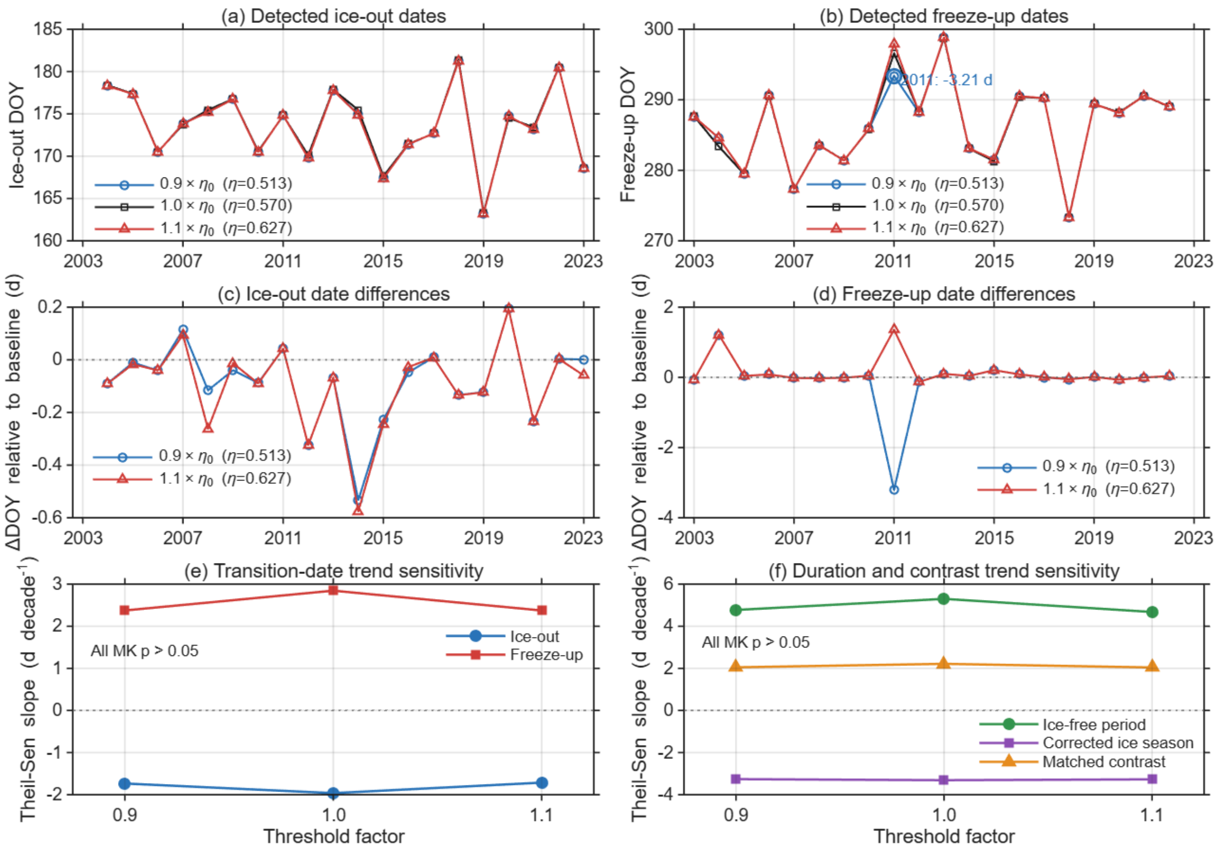
We evaluated the temporal-support assumptions using 18 prespecified complete detector configurations. Ice-out and freeze-up persistence requirements were varied independently among 3, 5, and 7 d, and the longer-term confirmation duration was set to either 10 or 14 equivalent observed days. The opposing trend directions were preserved across all configurations: stable ice-out slopes ranged from −1.97 to −1.36 d decade^{-1}, stable freeze-up slopes from +2.77 to +3.03 d decade^{-1}, ice-free-period slopes from +2.64 to +5.74 d decade^{-1}, corrected cross-year ice-season slopes from −3.86 to −2.70 d decade^{-1}, and matched-year trend-magnitude contrast slopes from +1.92 to +2.85 d decade^{-1}. None of the configuration-specific Mann–Kendall tests reached the 0.05 significance level.

We also recalculated the ±10 % state-threshold sensitivity test using the revised symmetric detector. The tested thresholds were η = 0.513, 0.570, and 0.627, corresponding to factors of 0.9, 1.0, and 1.1 relative to the baseline. All other detector settings were held constant.

Annual ice-out dates were only weakly affected: the median absolute difference relative to the baseline was 0.09 d for both the lower and upper thresholds. The corresponding maximum differences were 0.53 and 0.58 d, respectively, both occurring in 2014, and no ice-out date changed by more than 1 d. For freeze-up, the median absolute difference was 0.05 d for both thresholds. The largest sensitivity occurred in 2011, when the lower threshold advanced freeze-up by 3.21 d and the upper threshold delayed it by 1.36 d.

Despite this event-level exception, the trend directions remained unchanged. Across the three thresholds, Theil–Sen slopes ranged from −1.96 to −1.71 d decade^{-1} for ice-out, from +2.37 to +2.84 d decade^{-1} for freeze-up, from +4.68 to +5.30 d decade^{-1} for the ice-free period, from −3.31 to −3.26 d decade^{-1} for corrected ice-season duration, and from +2.04 to +2.21 d decade^{-1} for the matched-year trend-magnitude contrast. None of the corresponding

Mann–Kendall tests was statistically significant (all p > 0.05). The annual date comparisons and trend sensitivities are summarized in Response Figure R1-1.



Response Figure R1-1. Sensitivity of detected sea-ice transition dates and trends to a ±10 % perturbation of the AIF state threshold. The tested thresholds were η = 0.513, 0.570, and 0.627. Panels (a–d) show the annual transition dates and their differences from the baseline threshold, and panels (e–f) show the corresponding Theil–Sen slopes.

Finally, we formally tested whether the delay in freeze-up differed in magnitude from the advance in ice-out. We constructed a matched-year trend-magnitude contrast for the 19 years containing both stable boundaries, with a positive slope indicating that the freeze-up delay exceeded the magnitude of the ice-out advance. Its Theil–Sen slope was +2.21 d $decade^{-1}$, with a 95 % confidence interval from −4.13 to +7.77 d $decade^{-1}$ and a Mann–Kendall p value of 0.327. The complementary ordinary least-squares estimate was +2.06 d $decade^{-1}$, with a 95 % confidence interval from −4.10 to +8.22 d $decade^{-1}$ and p = 0.490.

These analyses show that the opposing directions of the estimated ice-out and freeze-up tendencies were preserved within the tested persistence, confirmation, and threshold ranges. However, although the point estimates suggest that the freeze-up delay was greater than the ice-out advance, the difference in trend magnitude was not statistically significant over the available record. We have therefore framed the result as a robust directional contrast within the tested detector settings, rather than as proof of a statistically significant difference or of physical causation by itself.

Changes to be incorporated in the revised manuscript: Equivalent direction-reversed early- and stable-transition criteria and the detector-configuration sensitivity analysis will be added to Sects. 3.3.2 and 3.3.3 and summarized in Table 2. The resulting transition estimates, sensitivity ranges, and formal matched-year trend-magnitude comparison are reported in Sect. 4.2.2, Table 4, Figure 9, and the Conclusions.

Major comment 3: Validation of the decadal inter-calibration

Status: Addressed

Reviewer comment. *Bridging the 2015-12-16 receiver upgrade is the core enabling claim, yet its validation is thin. How is the ~27% magnitude reduction derived, and is the assumption that the hardware change scales the ice signal and the summer baseline proportionally (so the ratio normalization cancels it) tested rather than asserted? An explicit homogeneity test showing no step change in the phenology residuals across the upgrade boundary would help resolve this question. For example, that pre-2016 and post-2015 transition dates are not systematically offset once calibrated.*

Response. We agree that the original explanation of the receiver replacement and the reported amplitude reduction was insufficient. In particular, the original manuscript did not clearly describe how the approximately 27 % difference was calculated and could be read as assuming that the receiver replacement proportionally scaled the winter ice signal and summer reference level. We have therefore re-examined the station history, clarified the normalization method, recalculated the amplitude difference, and tested directly for a remaining discontinuity in the transition dates.

The official IGS site log shows that the ASHTECH UZ-12 receiver was replaced by a TRIMBLE NETR9 on 16 December 2015. The antenna, radome configuration, cable, monument geometry, and marker eccentricities remained unchanged. The changes in July 2005 and July 2022 were firmware updates without receiver replacement. Thus, the December 2015 event was the only receiver-hardware replacement affecting the analysed AIF record.

We recalculated the amplitude difference from the processed AIF series. For each year, we first calculated the November–December mean of the reference-centred, low-pass-filtered AIF and then averaged these annual values with equal weight. The mean was 4.36 ± 0.47 for the 12 pre-replacement years from 2003 to 2014 and 3.25 ± 0.50 for the seven post-replacement years from 2016 to 2022, where the values following ± are interannual standard deviations. The post-replacement-period mean was therefore 25.6 % lower than the pre-replacement-period mean. The mixed-receiver year 2015 was excluded, and 2023 was unavailable because the record ended before November. Repeating the calculation before 10 min resampling produced comparable reductions of 25.7 % using filtered source-timestamp observations and 25.1 % using annual averages of daily medians. The difference was therefore not introduced by fixed-grid interpolation. However, because the two hardware periods contain different years and environmental conditions, we no longer attribute the complete 25.6 % difference exclusively to the receiver replacement.

We also clarified that the normalization is additive rather than ratio based. For each satellite and year, the mean AIF during the DOY 200–250 open-water reference interval is subtracted from the corresponding observations. This procedure aligns the annual open-water reference level and reduces satellite-dependent and interannual baseline offsets, but it does not assume that the winter ice signal and summer open-water baseline are scaled by the same proportion. Seasonal AIF amplitudes are therefore allowed to differ among satellites, years, and hardware periods.

Finally, we conducted two direct tests for a remaining step in the stable transition dates. After removing the full-record Theil–Sen trends, two-sided Mann–Whitney tests comparing the pre- and post-replacement residuals gave $p = 0.728$ for stable ice-out and $p = 0.874$ for stable freeze-up. Complementary segmented regressions estimated a post-replacement step of +1.98 d for ice-out, with a 95 % confidence interval from −6.50 to +10.47 d and $p = 0.628$, and −2.57 d for freeze-up, with a 95 % confidence interval from −13.75 to +8.61 d and $p = 0.634$.

These tests provide no evidence of a statistically detectable discontinuity in the stable transition dates across the December 2015 receiver replacement. We have nevertheless stated explicitly that the absence of a detectable phenological step does not demonstrate complete homogeneity of the AIF amplitudes or prove that all receiver-related effects were removed.

Changes to be incorporated in the revised manuscript: The receiver and firmware history has been clarified in Sect. 2.2. The additive reference-centring procedure, calculation of the 25.6 % period difference, resampling sensitivity checks, and transition-date homogeneity tests will be added to Sect. 3.3.1. The corresponding results and limitations are reported in Sect. 4.2.1 and the Conclusions.

Major comment 4: Selection of the degree-day windows

Status: Addressed

Reviewer comment. *The FDD window for freeze-up ($R^2 = 0.74$) and the TDD window for ice-out ($r = -0.81$) are chosen by a 2-D search over accumulation-window centre and duration to maximise correlation (Fig. 10, "white stars denote maximum negative correlations"). The reported $p < 0.001$ is not corrected for that search, so the couplings look more robust than they are. Cross-validation, out-of-sample test, or a multiplecomparison correction may help back up these claims and enable presenting these as established thermodynamic couplings.*

Response. We agree that the nominal p values in the original submission did not account for selection of the accumulation windows and therefore overstated the statistical evidence. We have replaced them with window-search-adjusted p values obtained using a max-statistic permutation test.

Candidate accumulation periods of 30, 60, 90, 120, 150, and 180 d were shifted in 1 d increments over the year. Ice-out and freeze-up were analysed separately, and the selected window was the one producing the maximum absolute Pearson correlation between the accumulated degree days and the corresponding stable transition dates.

In each of 9,999 permutations, the annual phenological dates were randomly reassigned among years, the complete search over all window positions and durations was repeated, and the maximum absolute correlation was retained. The adjusted p value was calculated from the proportion of permutation maxima at least as large as the observed maximum. Each candidate window retained its own valid-year mask. This procedure directly accounts for selection of the strongest relationship from the complete two-dimensional window search.

After correction, the selected TDD window for stable ice-out was DOY 84–173, corresponding to 90 d. The relationship had a Pearson correlation coefficient of −0.817, $R^2 = 0.668$, and a window-search-adjusted p value of 0.0004. The selected FDD window for stable freeze-up was DOY 260–289, corresponding to 30 d, with $r = -0.877$, $R^2 = 0.769$, and an adjusted p value of 0.0001. The revised Figure 11, reproduced below, presents both the complete window searches and the relationships for the selected accumulation periods.

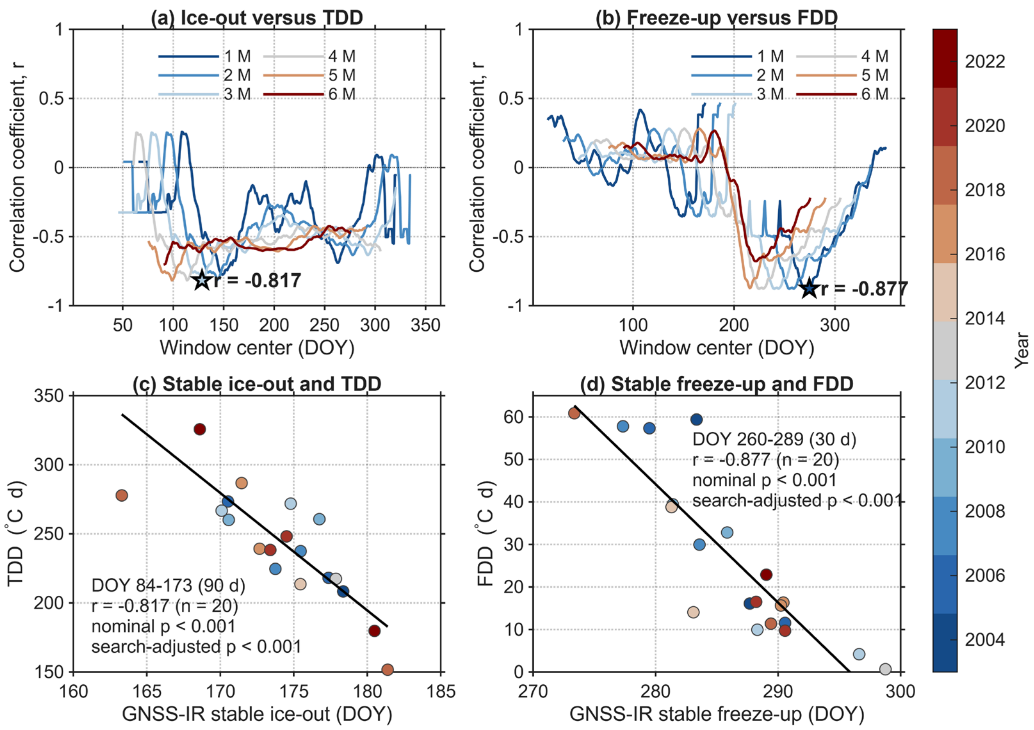


Figure 11. Sliding-window relationships between GNSS-IR stable-boundary phenological dates and accumulated thermal forcing. (a–b) Pearson correlation coefficients as functions of accumulation-window position and duration for (a) ice-out versus TDD and (b) freeze-up versus FDD; stars indicate the selected windows. (c) Relationship between stable ice-out and TDD accumulated over DOY 84–173. (d) Relationship between stable freeze-up and FDD accumulated over DOY 260–289. Reported p values for the selected relationships are adjusted for the complete window search.

We have also moderated the interpretation. The revised manuscript describes the results as physically consistent interannual associations rather than proof that the selected periods are unique or that temperature exclusively controls either transition. Although the absolute FDD correlation was greater than the TDD correlation, this difference was not formally tested and is not interpreted as evidence that freeze-up was significantly more temperature controlled than ice-out.

Changes to be incorporated in the revised manuscript: The max-statistic permutation procedure and complete accumulation-window search will be added to Sect. 3.3.4. The search-adjusted relationships and their revised interpretation are reported in Sect. 4.3.2. Figure 11 and its caption will be revised to show the selected windows and identify the reported p values as adjusted for the complete window search.

Major comment 5: Frazil-detection interpretation and the ice-out comparison

Status: Addressed

Reviewer comment. *The claim that GNSS-IR resolves "initial nearshore frazil ice formation" rests on a single freeze-up event (2019) with Sentinel-2 imagery, then is generalised across two decades. Meanwhile the ice-out comparison against IMS shows little interannual correlation (r = 0.05, p = 0.85), which raises the question of how the ice-out retrieval is validated across the record. The paper also offers two distinct explanations for the 5.5-day freeze-up lead — spatial dilution of the coarse IMS pixel, and earlier freezing of shallow nearshore water — without disentangling them. Additional cloud-free transition cases and a clearer separation of these mechanisms would substantially strengthen the central claim.*

Response. We agree that the original interpretation exceeded the available evidence. Selected optical satellite images from one year, comprising three Sentinel-2A scenes and two Landsat 8 scenes, cannot identify the ice type or formation process, explain the multi-year GNSS-IR–IMS timing difference, or support a record-wide claim of frazil-ice detection. We have therefore removed the claim that GNSS-IR detects "initial nearshore frazil ice formation." The five available scenes from 8 and 9 June and 15, 16, and 20 October 2019 are will used only as qualitative spatial context for that individual transition season; no quantitative ice fractions or ice-type classifications are inferred from them.

We also rechecked the IMS extraction code and corrected an inconsistent description of its spatial support. The comparison used four fixed neighbouring 4 km IMS grid cells surrounding TUKT. Daily IMS ice fraction was calculated as the proportion of these cells classified as ice, and the neighbourhood was considered ice covered when at least two of the four cells indicated ice. The cells have a nominal combined area of approximately 64 km^2 within an approximately 12 km × 8 km envelope. References in the original submission to a single central pixel and an 11 × 11-pixel window were reporting errors; the underlying IMS time series has not been changed.

Using the corrected stable-transition series, the 19 common freeze-up dates from 2004 to 2022 had a Pearson correlation coefficient of 0.845 ($p < 0.001$). The mean GNSS-IR-minus-IMS difference was −3.22 d, with a standard deviation of 3.41 d and an RMSE of 4.63 d; GNSS-IR detected freeze-up earlier in 15 of the 19 years. Differences in spatial support, sensitivity to partial or newly formed ice cover, and actual nearshore spatial variability may all contribute to this difference. However, the available data cannot separate or quantify their individual contributions, and the revised manuscript no longer attributes the difference solely to either IMS spatial dilution or earlier freezing of shallow nearshore water.

The weak ice-out comparison is will reported explicitly. Stable ice-out had $r = 0.048$ and $p = 0.846$ relative to IMS. The mean GNSS-IR-minus-IMS difference was −0.58 d, but its standard deviation was 7.45 d, the RMSE was 7.27 d, and only 8 of the 19 annual differences were within ±5 d. Thus, the small mean difference results from compensation between positive and negative annual differences and is not interpreted as close year-to-year agreement. This weak correspondence is consistent with the different spatial supports and the greater spatial and temporal complexity of coastal breakup, but it does not provide strong external validation of the individual GNSS-IR ice-out dates.

To assess the physical basis of AIF across the record, we compared observed elevation-dependent wavelet-power profiles for stable ice and open water. The analysis included 18 complete years and 637 matched year–PRN–azimuth groups. The elevation-integrated stable-ice-minus-open-water difference was positive in all 18 years, with a two-sided paired sign-flip p value below 0.001. This result supports the observed ice–water contrast represented by AIF, but it is not presented as independent validation of individual transition dates. The revised Figure 4, reproduced below, shows this observed contrast.

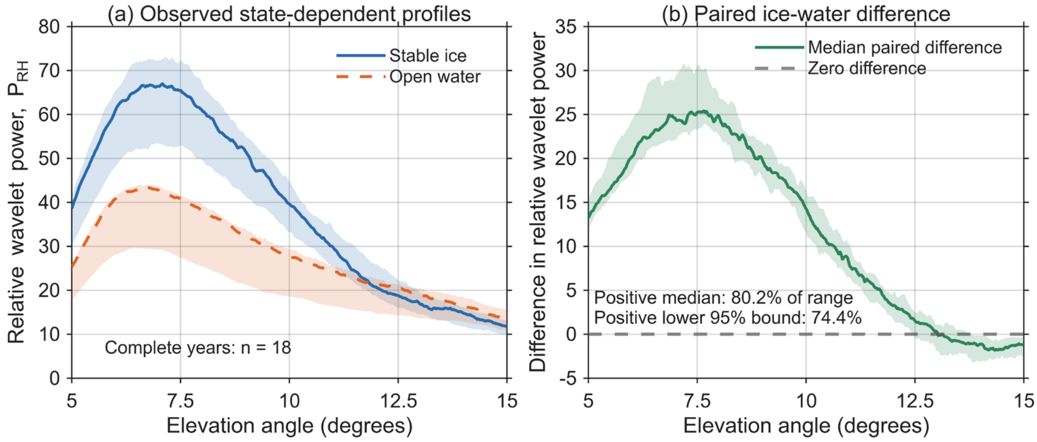
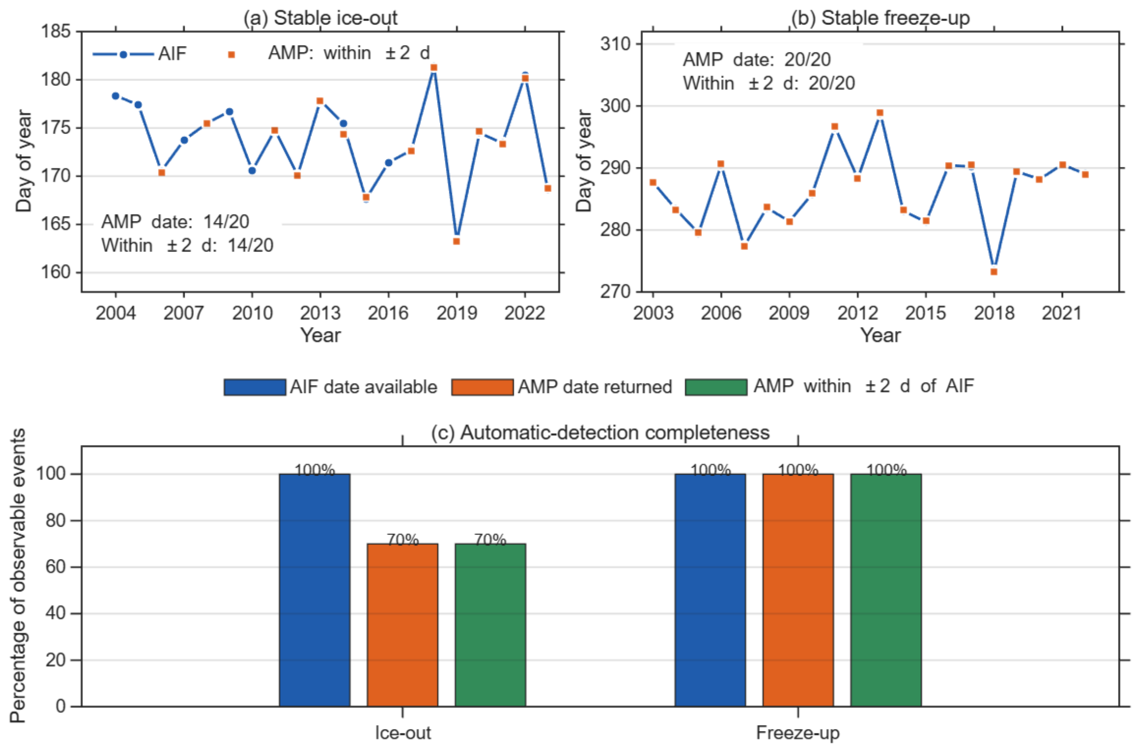


Figure 4. Observed elevation-dependent relative wavelet-power profiles for stable ice and open water. (a) Median profiles across 18 complete and equally weighted observation years from 2004 to 2022, excluding the mixed-receiver year 2015. (b) Median paired ice-minus-water difference. Shading denotes pointwise 95 % bootstrap intervals obtained by resampling the annual profiles. Profiles were matched within year, PRN, and azimuth cluster, yielding 637 matched groups. The gray dashed line in panel (b) denotes zero difference.

We additionally compared AIF with the conventional dominant Lomb–Scargle peak amplitude after applying the same reference centring, 24 h moving-median preprocessing, filtering, and rule-constrained detection framework. AIF returned all 20 observable stable ice-out dates, whereas peak amplitude returned 14. All 14 amplitude-derived dates were within 2 d of the corresponding AIF dates, with a mean amplitude-minus-AIF difference of −0.12 d, an RMSE of 0.33 d, and a maximum absolute difference of 1.15 d. The difference in event-return completeness had an exact McNemar p value of 0.031. For freeze-up, both indicators returned all 20 observable dates; the mean difference was +0.06 d, the RMSE was 0.12 d, and the maximum absolute difference was 0.34 d.

This comparison shows that conventional peak amplitude closely reproduces the transition timing when it returns a stable event, whereas AIF provides greater automatic detection completeness for ice-out. Because both indicators are derived from the same GNSS observations, this is an internal method comparison rather than independent external validation. The annual comparisons and event-return completeness are summarized in Response Figure R1-2.



Response Figure R1-2. Comparison of stable transition dates and automatic event-return completeness for AIF and conventional dominant Lomb–Scargle peak amplitude after the same DOY 200–250 reference centring, 24 h moving-median preprocessing, filtering, and rule-constrained detection. This internal comparison is provided in response to the reviewer and is not treated as independent validation.

Accordingly, the revised manuscript distinguishes three levels of evidence: The 2019 Sentinel-2A and Landsat 8 imagery provides qualitative spatial context for the individual transition season; IMS provides an external but spatially mismatched comparison of annual transition timing; and the observed-profile and AIF–amplitude analyses provide record-wide physical and internal methodological support. None is used to claim direct identification of frazil ice or complete external validation of every transition date.

Changes to be incorporated in the revised manuscript: The frazil-ice claim will be removed from the Abstract, Results and discussion, and Conclusions. The IMS spatial support has been corrected and unified in Sect. 2.3.2, and the IMS comparison and its limitations will be revised in Sect. 4.4. The observed-profile and AIF–amplitude comparison methods and results will be added to Sects. 3.2, 3.3.2, and 4.1. Figure 4 and its caption will be revised. Response Figure R1-2 is provided here as an additional internal method comparison and will not be included in the revised manuscript.

Specific Comments

Specific Comments 1: *Abstract-"statistically significant shortening … 4.63 days per decade (p = 0.04)". The abstract does not state that this is the corrected-duration metric with 2018 excluded. State the metric explicitly and give the with-2018 value (p = 0.06) alongside it, so the headline is not read as an unconditional result.*

Response. We agree that the original Abstract presented this result without adequately identifying the metric or explaining the exclusion of 2018. Rather than retaining the former comparison with and without 2018, we have removed that exclusion from the primary analysis. All years with valid observations, including 2018, are will retained, and no year is excluded because of its effect on the fitted slope or p value.

The primary analysis will use Theil–Sen slopes and two-sided Mann–Kendall tests. Under the primary stable–stable definition, corrected cross-year ice-season duration tended to decrease at −3.31 d decade^{-1}, with a 95 % confidence interval from −7.82 to +0.63 d decade^{-1} and $p = 0.086$. The original estimates based on excluding 2018 (4.63 d decade^{-1}; $p = 0.04$) and retaining it in the former ordinary least-squares analysis ($p = 0.06$) are therefore no longer used as primary evidence. The revised Abstract identifies the metric explicitly and describes the result as a directional tendency rather than a statistically significant change. The complete statistical revision is described in our response to Major Comment 1.

Changes to be incorporated in the revised manuscript: The original significance-led statement has been replaced in the Abstract. The revised statistical methods and complete results are reported in Sects. 3.3.3 and 4.2.2, Table 4, Figure 9, and the Conclusions.

Specific Comments 2: *Introduction- the context is thorough, but the framing is often heavily promotional. For TC I would tone the language to match the strength of the evidence given n ≈ 19 and the wide confidence intervals. It leans heavily on intensifiers and evaluative adjectives (examples throughout the text - "profound," "extreme," "rigorously," "true," "genuine", "anomalously", "exceptionally", "exact", "clearly", "remarkable", "crucially", "disruptive", "resounding", "ultimate", "conclusively", "robust", "highly reliable") that overstate. Given the modest sample size and wide confidence intervals, this undercuts credibility. The results would read as more authoritative in plain concise language using figures and statistics in place of adjectives.*

Response. We agree that the language in the original submission was frequently stronger than warranted by the statistical and observational evidence. We therefore conducted a manuscript-wide revision to remove promotional intensifiers and replace them with quantitative estimates, uncertainty ranges, and evidence-based interpretation.

Definitive statements such as "conclusively validates" have been replaced by more specific formulations such as "supports the physical consistency of the record." Physical explanations that cannot be separated using the available observations are now expressed as possible interpretations rather than established mechanisms. We also distinguish consistently among directional tendencies, statistically significant trends, sensitivity results, and qualitative supporting observations.

The original title referred to an "asymmetric decline," although the formal comparison did not establish a statistically significant difference between the ice-out and freeze-up trend magnitudes. The title has therefore been changed to "Two decades of GNSS-IR observations of coastal sea-ice phenology indicate contrasting ice-out and freeze-up tendencies at Tuktoyaktuk, Beaufort Sea." This wording retains the observed directional contrast without implying that a statistically significant asymmetry has been demonstrated.

Changes to be incorporated in the revised manuscript: The Title, Abstract, Introduction, Results and discussion, figure captions, and Conclusions will be revised. Descriptive intensifiers will be removed or replaced with numerical estimates, uncertainty ranges, and explicitly qualified interpretations.

Specific Comments 3: *Introduction – the review of GNSS-R sea-ice work is almost entirely ground-based / near-nadir. A short acknowledgement of the broader (including spaceborne) GNSS-R sea-ice literature would better position the specific contribution of a fixed ground station, and clarify what ground-based GNSS-IR adds that the spaceborne approaches cannot (fixed footprint, high temporal continuity at one nearshore point).*

Response. We agree that the original Introduction did not adequately represent the broader development of GNSS-R sea-ice remote sensing. We have expanded the literature review to include representative spaceborne applications, including TechDemoSat-1 sea-ice detection and classification, spaceborne phase altimetry over sea ice, and more recent retrievals of Arctic sea-ice concentration, thickness, and related physical properties (Li et al., 2017; Rodriguez-Alvarez et al., 2019; Cartwright et al., 2019; Yang et al., 2023; Xie and Yan, 2024; Muñoz-Martín et al., 2025).

We have also clarified the complementary roles of the observing platforms. Spaceborne GNSS-R provides broad regional sampling across polar areas, whereas a fixed ground-based GNSS-IR station repeatedly observes a defined nearshore sector and can provide high temporal continuity over multiple years. The present study is therefore positioned as a site-scale, temporally continuous complement to regional satellite observations rather than as a replacement for spaceborne GNSS-R.

Changes to be incorporated in the revised manuscript: The GNSS-R literature review in the Introduction has been expanded to include representative spaceborne sea-ice applications and to explain the complementary spatial and temporal roles of spaceborne GNSS-R and fixed ground-based GNSS-IR observations. Six relevant references will be added to this discussion.

Specific Comments 4: *Section 2.2 / 2.3 – data completeness is quoted as 97.5% in the text but 97% in Table 1; reconcile.*

Response. Thank you for identifying this inconsistency. The value has been standardized to 97.5 % in both the text and Table 1, and its denominator has been clarified.

The original daily-file archive spans 21 August 2003 to 6 September 2023 and contains 7,136 available GNSS daily files out of 7,322 calendar days, corresponding to 97.5 % daily-file availability. Following AIF processing, original AIF observations were available on 7,135 days. These quantities describe archive availability and AIF observation coverage, respectively.

We have also clarified that the 97.5 % value does not represent satellite-arc retention after quality control. The final AIF series comprised 135,975 retained arcs. The conditional signal-quality retention rate was 90.0 % among spectrally eligible arcs, whereas the end-to-end retention rate was 72.1 % relative to all initially segmented candidate arcs.

Changes to be incorporated in the revised manuscript: The completeness value has been standardized to 97.5 % in Sect. 2.2 and Table 1. Sect. 2.2 will distinguish daily-file availability, AIF observation coverage, and satellite-arc retention after quality control.

Specific Comments 5: *Section 3.1 (physical basis; Fig. 2, Fig. 3) — The technique operates at 5–15° elevation, i.e. 75–85° incidence, which is a grazing-angle geometry, but this is never stated and the term "grazing" does not appear. Figure 2 compounds this by depicting satellites high overhead with steep reflections, which misrepresents the near-horizon geometry used; a schematic drawn at representative low elevation would serve the reader better. More substantively, the physical mechanism in Section 3.1.2 is mis-attributed. Seawater (ε ≈ 70) is a far stronger reflector than sea ice (ε ≈ 3) on dielectric contrast alone, so "stark dielectric contrast" as written predicts the opposite of the ordering shown in Fig. 3 (green above red). The ordering in Fig. 3 is in fact defensible, but for a reason the manuscript does not give: at grazing incidence the relevant quantity is the co-polar RHCP reflection coefficient, whose angular behaviour is set by the pseudo-Brewster angle. Seawater's pseudo-Brewster (~83° incidence, ~7° elevation) falls inside the 5–15° operating window, suppressing the RHCP-reflected component from open water there, whereas sea ice's (~60° incidence, ~30° elevation) does not — so ice retains the higher RHCP coefficient across the operational range. This grazing-angle polarization effect, not dielectric contrast, is what underpins the AIF's ice/water discrimination, and it should be explained as such.*

Response. We agree with both the geometric and electromagnetic aspects of this comment. The revised manuscript now states explicitly that the selected satellite elevation interval of 5°–15° corresponds to incidence angles of approximately 75°–85° and therefore represents grazing-angle reflection geometry. Figure 2 has been redrawn to show representative low-elevation satellite paths and shallow reflections rather than high-elevation satellites with steep reflection paths.

We also agree that bulk dielectric contrast alone does not explain the ordering of the modelled sea-ice and seawater responses. Under grazing geometry, the signal received by the RHCP geodetic antenna depends on the circular-polarization Fresnel response, together with antenna response and surface roughness. Under the assumptions used in the theoretical calculation, the pseudo-Brewster region of seawater lies within or close to the 5°–15° operating range and suppresses the co-polar RHCP component from open water. The corresponding pseudo-Brewster region for sea ice lies outside this range, allowing sea ice to retain a stronger modelled co-polar RHCP response over much of the operational interval.

Section 3.1.2 has therefore been rewritten around the grazing-angle co-polar RHCP response rather than a general dielectric-contrast argument. Figure 3 will distinguish explicitly among the antenna response, surface-roughness attenuation, and co-polar RHCP Fresnel terms. The revised Figures 2 and 3 are reproduced below.

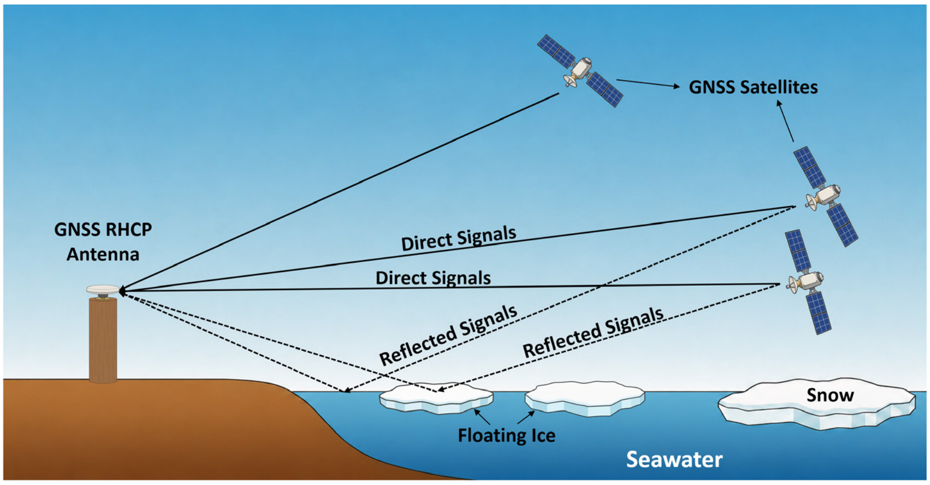


Figure 2. Schematic of low-elevation GNSS-IR observations over the nearshore surface. The direct and reflected signals are received simultaneously by the geodetic antenna, and their elevation-dependent interference produces oscillations in the recorded SNR.

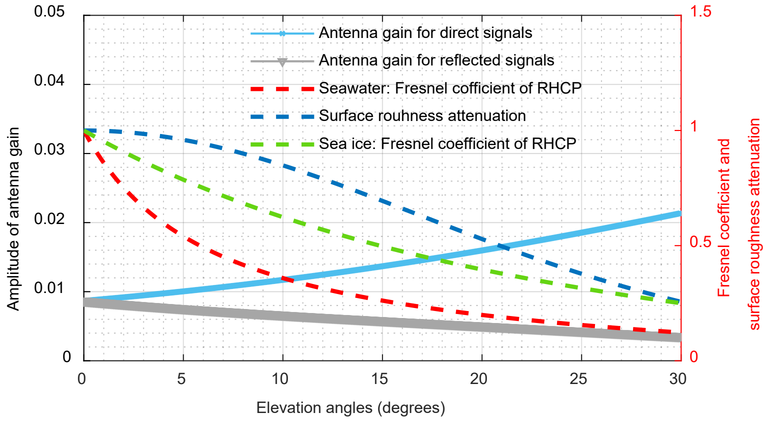


Figure 3. Decomposition of factors affecting the reflected-signal amplitude. Solid lines show the antenna responses, the blue dashed line shows surface-roughness attenuation, and the green and red dashed lines show the co-polar RHCP Fresnel terms for sea ice ($\varepsilon_{\text{ice}} \approx$ 3.0 + 0.004i) and seawater (ε_{sw} = 70 + 64i), respectively. The selected operating range is 5°–15°.

As an observational check on this theoretical expectation, stable ice produced a greater elevation-integrated wavelet-power response than open water in all 18 complete comparison years, with a paired sign-flip p value below 0.001. The contrast weakened above approximately 13°, indicating that the observed ice–water separation is elevation dependent and that the theoretical model should not be interpreted as a complete description of all surface and propagation effects. This observational comparison is shown in the revised Figure 4 and is discussed further in our response to Major Comment 5.

Changes to be incorporated in the revised manuscript: The grazing-angle geometry and corresponding incidence-angle range are now stated in Sect. 3.1.1. The physical explanation in Sect. 3.1.2 has been rewritten in terms of the co-polar RHCP Fresnel response and pseudo-Brewster behaviour. Figure 2 has been redrawn, Figure 3 and its caption will be revised, and the record-wide observed state comparison will be added to Sect. 4.1 and Figure 4.

Specific Comments 6: *Section 3.1.2 – the reflectivity numbers are internally inconsistent. "At a 5° elevation angle, the reflectivity for ice is ~23 versus ~0.22 for seawater, representing a ~45% increase" — 23 vs 0.22 is a factor of ~100, not a 45% increase, and it does not match the "factor of 2–3" amplitude ratio stated in the next sentence. Please correct these figures and make the quantities (permittivity vs reflection coefficient vs amplitude) explicit.*

Response. We agree that the numerical comparison in the original submission was incorrect and that the quantities were not defined consistently. The values of approximately 23 and 0.22, the stated 45 % increase, and the subsequent factor-of-two-to-three amplitude ratio have therefore been removed.

We recalculated the relevant electromagnetic quantity using the assumptions adopted for Figure 3. Averaged over the operational elevation range of 5°–15°, the magnitude of the modelled co-polar RHCP Fresnel term was approximately 0.78 for sea ice and 0.54 for seawater. The modelled sea-ice value was therefore approximately 44 % greater over this range.

The revised text will distinguish explicitly among complex permittivity, which is an input describing the electromagnetic properties of the medium; the co-polar RHCP Fresnel term, which is a modelled reflection-response magnitude; the observed reflected-signal amplitude, which also depends on antenna response, surface roughness, and propagation effects; and AIF, which is an elevation-integrated relative wavelet-power metric. The modelled Fresnel difference is used only to explain the expected direction of the ice–water contrast and is not interpreted as a direct prediction of the measured AIF ratio.

Changes to be incorporated in the revised manuscript: The inconsistent reflectivity values and unsupported amplitude-ratio statement will be removed from Sect. 3.1.2. The corrected mean co-polar RHCP Fresnel responses and the distinctions among permittivity, Fresnel response, reflected-signal amplitude, and AIF are now stated explicitly in Sect. 3.1.2 and Figure 3. The revised Figure 3 is reproduced in our response to Specific Comment 5.

Specific Comments 7: *Section 3.2 – please clarify a few AIF definitions for reproducibility: what σ_Rh = 0.2 m represents the choice of n = 800 interpolated elevation samples, and the physical units of Λ (a dimensionless "calibrated AIF"?). These would help future users reuse the metric.*

Response. We agree that these definitions were not sufficiently clear for reproducing the AIF calculation. The value σRh = 0.2 m represents an assumed one-standard-deviation uncertainty in the nominal antenna-to-reflector height. It is used only to define the wavelet-scale averaging interval, extending from the nominal reflector height minus three standard deviations to the nominal height plus three standard deviations. It is not an uncertainty estimate for AIF or for the detected transition dates, and the same value is applied throughout the record.

For numerical integration, the elevation-dependent wavelet-power curve is interpolated to 800 uniformly spaced samples between 5° and 15°, corresponding to an approximate spacing of 0.0125°. These points standardize the numerical integration across satellite arcs with different original sampling patterns. They are not independent GNSS observations and do not increase the effective sample size.

Finally, the processed wavelet power is expressed in relative units. Because the elevation integration includes conversion from degrees to radians, Λ and its annual reference-centred form are dimensionless relative measures. They are not radiometrically calibrated power values, and the term "calibrated AIF" has been replaced by "reference-centred AIF" where appropriate.

Changes to be incorporated in the revised manuscript: The meaning and role of σRh = 0.2 m, the construction and purpose of the 800-point elevation grid, the degree-to-radian conversion, and the dimensionless relative nature of Λ have been clarified in Sect. 3.2 immediately following the AIF equations.

Specific Comments 8: *Section 3.3.1 – the QC retains ~90% of observations. How do the discarded 10% differ from the retained set — are they biased by season (e.g. are noisier summer tracks preferentially removed, which would affect the ice-free baseline), by satellite, or across the pre/post-2015 boundary? A short characterisation would reassure that screening does not itself shape the baseline or the trend.*

Response. We agree that the original retention percentage was insufficiently defined. We therefore conducted a complete arc-level quality-control audit and clarified both the denominator and the principal causes of exclusion.

Across the record, 188,473 segmented candidate arcs were evaluated. After the minimum-sample, maximum-duration, uniform-grid, and complete 5°–15° elevation-span checks, 151,127 arcs were eligible for spectral quality assessment. Of these, 135,975 passed the reflector-height, peak-to-noise, and maximum-amplitude criteria. The previously reported retention of approximately 90.0 % therefore refers to the spectrally eligible arcs. Relative to all initially segmented candidates, the end-to-end retention rate was 72.1 %. The largest individual exclusion sources were incomplete elevation-span coverage, which removed 24,575 arcs, and insufficient samples, which removed 12,471 arcs.

Conditional signal-quality retention was lowest during March–May, at 77.7 %–84.8 %, and highest during November–January, at 97.7 %–99.6 %. During the DOY 200–250 reference interval, conditional retention was 86.6 %, compared with 90.6 % outside this interval; the corresponding Cramér's V was 0.048. End-to-end retention was 73.6 % within the reference interval and 71.9 % outside it. Thus, although signal-quality retention was seasonally variable, the complete processing chain did not preferentially deplete observations within the summer reference interval.

Conditional retention was 91.2 % before and 88.5 % after the receiver replacement, with Cramér's V = 0.045. The corresponding end-to-end rates were 66.2 % and 80.8 %, indicating that the larger receiver-period difference arose primarily during the preceding sampling and geometric-eligibility stages rather than the subsequent signal-quality screening. Conditional retention among individual satellites ranged from 79.5 % to 97.6 %, with Cramér's V = 0.115.

These results show measurable seasonal and satellite dependence, but only small associations between conditional signal-quality retention and either reference-period membership or receiver period. We now acknowledge that sampling and screening are not completely uniform and avoid claiming that quality control removes all related effects.

Changes to be incorporated in the revised manuscript: The quality-control denominator, stage-by-stage arc counts, principal rejection causes, and retention statistics stratified by season, reference interval, receiver period, and satellite will be added to Sect. 3.3.1. The final number of 135,975 retained arcs has also been reported in Sect. 2.2.

Specific Comments 9: *Section 3.3.1 – the reference-period normalization sets DOY 200–250 to zero-mean as the ice-free baseline. Please confirm this window is genuinely ice-free in every year of the record (it appears safe given a mean ice-out of DOY 173 and freeze-up of DOY 284, but this should be stated and evidenced) and note how many tracks/satellites typically contribute per day to that baseline.*

Response. We agree that the ice-free status and observational support of the reference interval should be demonstrated rather than assumed. We therefore checked DOY 200–250 against the final stable transition dates.

Across the complete transition series, the latest stable ice-out occurred on DOY 181.39 and the earliest stable freeze-up on DOY 273.36. Thus, DOY 200–250 followed stable ice-out and preceded stable freeze-up in every year with complete seasonal coverage. The endpoint years were only partly observed: the available 2003 observations occurred during open water before the detected freeze-up, whereas the available 2023 observations occurred after the detected ice-out and ended before freeze-up. These record limitations are now stated explicitly.

Across the 1,035 calendar days contained in the actually observed portions of DOY 200–250, every day included at least one retained satellite arc. A typical reference-period day contributed a median of 20 retained arcs from 20 unique satellites, with an interquartile range of 18–22 for both quantities. The reference mean is calculated separately for each satellite and year before observations are combined, thereby avoiding a single correction across satellites, years, or hardware periods.

Changes to be incorporated in the revised manuscript: The stable-transition bounds supporting selection of DOY 200–250, treatment of the two incomplete endpoint years, and daily arc and satellite counts will be added to Sect. 3.3.1 immediately after the definition of the open-water reference interval.

Specific Comments 10: *Section 3.3.2 – see key question 2 on the asymmetric detection rules. Extending the sensitivity analysis beyond ±10% on the threshold to include the persistence criterion would directly test whether the reported asymmetry survives.*

Response. We agree that varying only the state threshold did not test sensitivity to the asymmetric persistence requirements used in the original submission. We therefore replaced those rules with equivalent, direction-reversed criteria and added an explicit persistence and stability-confirmation sensitivity analysis.

Under the revised primary detector, stable ice-out and freeze-up use the same 7 d immediate support period, temporal-coverage requirement, longer-term stability confirmation, terminal seasonal anchor, and 5 d sustained-reversal test. The early-transition criteria are also symmetric, with a common 2 d persistence requirement for both ice-out and freeze-up.

We evaluated 18 complete prespecified configurations by varying ice-out and freeze-up persistence independently among 3, 5, and 7 d and the longer-term confirmation duration between 10 and 14 equivalent observed days. Stable ice-out slopes remained negative, ranging from −1.97 to −1.36 d decade^{-1}, whereas stable freeze-up slopes remained positive, ranging from +2.77 to +3.03 d decade^{-1}. The matched-year trend-magnitude contrast also remained positive, ranging from +1.92 to +2.85 d decade^{-1}. None of the configuration-specific Mann–Kendall tests reached the 0.05 significance level.

The original ±10 % threshold test was not discarded. It was recalculated using the revised symmetric detector and fixed 2 d persistence for both transitions. The opposing trend directions were also preserved across the three tested thresholds, although individual freeze-up dates, particularly in 2011, showed greater sensitivity. Full results of both sensitivity analyses are provided in our response to Major Comment 2 and in Response Figure R1-2.

Changes to be incorporated in the revised manuscript: The symmetric detector and persistence and confirmation-duration sensitivity design will be added to Sect. 3.3.2. The corresponding trend calculations and results are reported in Sects. 3.3.3 and 4.2.2, Table 4, Figure 9, and the Conclusions. The recalculated ±10 % threshold sensitivity analysis is documented in our response to Major Comment 2.

Specific Comments 11: *Section 4.1 (2019 case) – the five-phase interpretation is a nice narrative, but the physical attributions (e.g. the Phase II "rebound" as bare ice / melt ponds, Phase I decline as snow densification) are asserted with strong language for a single year. Please temper "physically represents" / "perfectly coincides" to reflect that these are plausible interpretations of one season's co-timing.*

Response. We agree that the original five-phase discussion assigned specific physical processes more confidently than the observations allowed. AIF responds to the combined effects of surface state, snow and ice properties, roughness, reflection geometry, and antenna response and therefore cannot independently distinguish all processes occurring during a single season.

We have rewritten the discussion to separate observations from possible explanations. For example, the Phase I decrease is now described as being consistent with progressive changes in the snow–ice surface, while acknowledging that the relative contributions of snow metamorphism, surface roughness, and volume scattering cannot be separated. The Phase II increase may reflect wet snow, exposed ice, or small meltwater features, but no single cause is assigned. Equivalent qualifications have been introduced for the later melt, open-water, and freeze-up phases.

Expressions such as "physically represents," "perfectly coincides," and "ultimate validation" will be removed. Meteorological data and Sentinel-2 imagery are will used only to provide temporal and spatial context. The 2019 record is presented as an illustrative example of seasonal AIF evolution rather than confirmation of the specific physical processes inferred for each phase.

Changes to be incorporated in the revised manuscript: The five-phase discussion in Sect. 4.1 and the interpretations associated with Figures 5–7 have been rewritten using conditional language and explicit observational limitations.

Specific Comments 12: *Section 4.2.2 (Table 4, Fig. 8) – the exclusion of 2018 is justified by the authors because it "dampens" the trend, which reads as motivated. The authors should cite precedent for excluding samples this way. It is stated to be "extreme noise" with no evidence. Define an objective, a-priori outlier criterion and present the with/without-2018 results even-handedly rather than foregrounding the version that reaches significance. Figure 8's caption says "2004 through 2022" while Table 4 and the text say 2004–2023 — reconcile the end year.*

Response. We agree that excluding 2018 because it reduced the fitted trend was not an objective basis for the primary analysis. In the revised manuscript, 2018 is retained together with all other years having valid observations. The former comparison with and without 2018 will be removed from the primary results, and 2018 is no longer described as "extreme noise." Because no year is now excluded as an outlier, an a-priori exclusion criterion is not required.

The primary analysis will use Theil–Sen slopes and two-sided Mann–Kendall tests, which are less sensitive than ordinary least-squares regression to individual influential years. Sensitivity is assessed through alternative early/stable boundary definitions and 18 prespecified detector configurations rather than by selectively removing annual observations.

We also reconciled the analysis periods, which differ because of the incomplete endpoint years. Stable ice-out covers 2004–2023 ($n = 20$), stable freeze-up covers 2003–2022 ($n = 20$), and the within-year ice-free period covers 2004–2022 ($n = 19$). Corrected cross-year ice-season duration covers 2004–2023 ($n = 20$) because each value links freeze-up in the preceding year with ice-out in the current year. These metric-specific periods and sample sizes are now stated explicitly.

Changes to be incorporated in the revised manuscript: The with/without-2018 analysis and the description of 2018 as "extreme noise" will be removed from Sect. 4.2.2. Metric-specific periods and sample sizes have been reconciled in Sect. 3.3.3, Table 4, the revised Figure 9, and their captions. The revised Figure 9 is reproduced in our response to Major Comment 1.

Specific Comments 13: *Section 4.4 (Table 5) – the ice-out row has essentially zero correlation with IMS ($r = 0.05$), yet ice-out trends are still reported and interpreted. Please discuss what, if anything, independently validates the ice-out dates across the 20-year record beyond the single 2019 Sentinel-2 case. Also the sign convention is confusing: the caption states negative = earlier detection by GNSS-IR, but the ice-out offset of −1.4 is labelled "(later)". Please make the convention consistent.*

Response. We agree that the weak ice-out correspondence with IMS and the limitations of external validation must be stated explicitly. No independent 20-year reference record exists with the same local spatial support as the GNSS-IR observations. The individual ice-out dates therefore cannot be externally validated on a year-by-year basis across the complete record.

In the revised comparison, stable ice-out had a Pearson correlation coefficient of 0.048 and p = 0.846 relative to IMS. The mean GNSS-IR-minus-IMS difference was −0.58 d, but its standard deviation was 7.45 d, the RMSE was 7.27 d, and only 8 of the 19 annual differences were within ±5 d. The small mean difference therefore results from compensation between positive and negative annual differences rather than close year-to-year agreement. Spatially heterogeneous breakup, wind-driven ice movement, fracturing, melt ponds, and differences in observational footprint are possible contributors, but their relative effects cannot be quantified using the available observations.

We will distinguish external validation from supporting evidence. The 2019 Sentinel-2 scenes provide qualitative event-scale context and are not treated as two-decade validation. Record-wide support is instead provided by the positive elevation-integrated stable-ice-minus-open-water contrast in all 18 complete comparison years and by the internal comparison with conventional peak amplitude. Peak amplitude returned 14 of the 20 observable stable ice-out dates, all within 2 d of the corresponding AIF dates, whereas AIF returned all 20. Because both indicators use the same GNSS observations, these results demonstrate physical consistency and internal reproducibility but do not constitute independent external validation.

The sign convention has also been standardized. All differences are defined as GNSS-IR minus IMS; consequently, a negative value always indicates earlier GNSS-IR detection. The incorrect “later” label associated with the negative ice-out difference will be removed. The former within-calendar-year ice-duration comparison has also been removed, leaving stable ice-out, stable freeze-up, and corrected cross-year ice-season duration in Table 5.

The revised Table 5 and Figure 12 use this convention consistently and are reproduced below.

Table 5. Statistical comparison between GNSS-IR and IMS phenological metrics. Stable ice-out and freeze-up comparisons cover 2004–2022, whereas the corrected cross-year ice-season-duration comparison covers 2005–2022. Bias is defined as GNSS-IR minus IMS.

Metric	Period and sample size	GNSS-IR minus IMS Bias ± SD	RMSE	Pearson correlation coefficient (p value)	Additional comparison
Stable ice-out date	2004–2022, n=19	−0.58±7.45d	7.27d	0.048 (0.846)	8/19 within ±5d
Stable freeze-up date	2004–2022, n=19	−3.22±3.41d	4.63d	0.845 (<0.001)	GNSS-IR earlier in 15/19 years
Corrected cross-year ice-season duration	2005–2022, n=18	+3.15±8.27d	8.64d	0.207 (0.411)	—

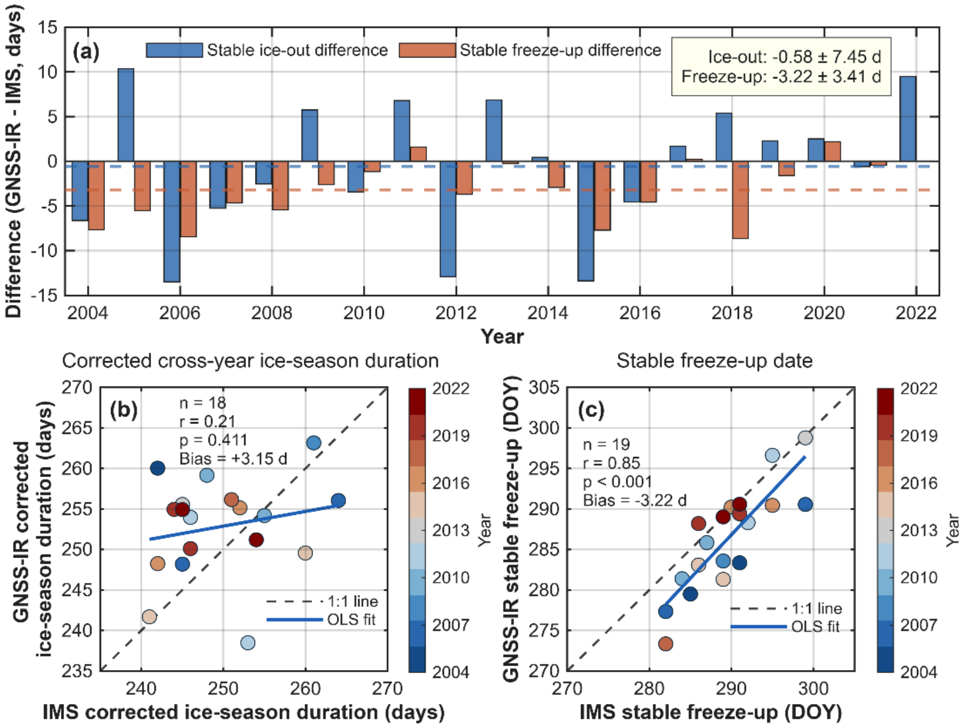


Figure 12. Cross-scale comparison of GNSS-IR and IMS sea-ice phenological metrics. (a) Annual differences in stable ice-out and freeze-up dates over 2004–2022, calculated as GNSS-IR minus IMS; negative values indicate earlier GNSS-IR detection, and the coloured dashed lines indicate the corresponding mean differences. (b) Relationship between corrected cross-year ice-season durations derived from IMS and GNSS-IR over 2005–2022. (c) Relationship between IMS and GNSS-IR stable freeze-up dates over 2004–2022. In panels (b) and (c), point colours indicate year, black dashed lines show the 1:1 relationship, and blue solid lines show the ordinary least-squares fits.

Changes to be incorporated in the revised manuscript: The limitations of external ice-out validation and the supporting internal analyses are now discussed in Sects. 4.1 and 4.4. The GNSS-IR-minus-IMS sign convention is defined in the comparison methods and used consistently in Table 5, Figure 12, and their captions.

Specific Comments 14: *Section 4.5 (Table 6) – the benchmarking against pan-Arctic passive-microwave trends is a useful addition, and the honest discussion of epoch and sensing-regime differences is appreciated. The agreement of the GNSS-IR trend (−4.19 ± 5.43) with the concurrent IMS trend (−4.23 ± 1.8) is described as “resounding cross-validation”, but the GNSS-IR interval is ~3× wider and IMS is not an independent ground truth at these scales. I would present the agreement as reassuring rather than conclusive.*

Response. We agree that the phrase “resounding cross-validation” was not justified. IMS is a neighbourhood-scale gridded product rather than an independent local ground-truth record, and the approximately two-decade GNSS-IR record has substantially greater trend uncertainty than longer regional satellite records. The phrase has therefore been removed.

Section 4.5 has been reframed as cross-scale trend benchmarking. Using the revised corrected cross-year definition and primary statistical framework, the full-record GNSS-IR trend was −3.31 d decade^{-1} for 2004–2023, with a 95 % confidence interval from −7.82 to +0.63 d decade^{-1} and $p = 0.086$. Over the common 2005–2022 period, the corresponding trends were −3.00 d decade^{-1} for GNSS-IR and −2.50 d decade^{-1} for IMS, with $p = 0.256$

and p = 0.424, respectively. Both datasets indicated a tendency toward a shorter continuous ice season, but neither trend was statistically significant.

Their common direction is described as broadly consistent rather than validating. Year-to-year agreement in corrected ice-season duration was weak (r = 0.207; p = 0.411), and the comparison is affected by differences in spatial support, observation period, sensing method, and transition definition. Published passive-microwave results are will used only to place the local estimate in a broader Arctic context. We explicitly state that the TUKT result should not be interpreted as a regional Beaufort Sea trend or as independent validation of satellite-derived estimates.

The revised Table 6 presents this comparison as contextual benchmarking and is reproduced below.

Table 6. Comparison of corrected ice-season trends in this study with published Arctic changes in ice-season, melt-season, and open-water-season duration. Values are expressed in days per decade. Negative corrected ice-season values indicate shortening, whereas positive melt- or open-water-season values indicate lengthening.

Study	Region	Period	Trend	Data/scale	Significance
This study	TUKT coast	2004–2023	−3.31	GNSS-IR, site-scale nearshore sector	Non-significant; p value 0.086
IMS	TUKT region	2005–2022	−2.50	IMS, four-cell neighbourhood (~64 km²)	Non-significant; p value 0.424
Stroeve et al., 2014	Pan-Arctic	1979–2013	~5	PMW, 25 km	Significant
Stroeve et al., 2014	Beaufort Sea	1979–2013	6–11	PMW, 25 km	Significant
Markus et al., 2009	Pan-Arctic	1979–2007	4.7–6.4	PMW, 25 km	Significant
Markus et al., 2009	Chukchi/Beaufort Seas	1979–2007	9.7–12	PMW, 25 km	Significant
Parkinson, 2014	Arctic	1979–2013	≈−5	PMW, 25 km	Significant
Johnson and Eicken, 2016	Beaufort coast	1979–2013	~15–19*	PMW, 25 km	Significant

*The value represents an approximate open-water-season lengthening inferred from the reported freeze-up delay and breakup advance rather than a directly reported corrected cross-year ice-season trend.

Changes to be incorporated in the revised manuscript: Sect. 4.5 and Table 6 will be revised to emphasize differences in scale, period, sensing method, metric definition, and statistical uncertainty. Claims of cross-validation have been replaced by cautious cross-scale and directional comparison.

Specific Comments 15: *Section 5 (Conclusions) – the statement that the method "requires no localized calibration" sits awkwardly next to Section 3.3, which is entirely an inter-calibration procedure. Please clarify what is meant (presumably: no external/auxiliary calibration data are needed because the summer baseline is self-referencing) so the two statements do not appear to contradict.*

Response. We agree that the statement that the method "requires no localized calibration" was ambiguous and inconsistent with the normalization procedure in Sect. 3.3.1. The method does require station-specific reference normalization, although it does not require an external radiometric calibration target.

The ambiguous statement has therefore been removed. The method is now described as internally referenced: for each satellite and year, the mean AIF during the DOY 200–250 open-water interval is subtracted from the corresponding observations. This additive reference centring aligns the annual open-water baseline before observations from different satellites, years, and hardware periods are combined.

This distinction also clarifies the limitations of the procedure. Reference centring does not assume proportional scaling of the winter ice signal and summer baseline, does not eliminate all differences in seasonal AIF amplitude, and does not demonstrate complete hardware homogeneity. It instead provides a consistent annual reference for applying the same state-classification and transition-detection framework across the record.

Changes to be incorporated in the revised manuscript: The statement that the method "requires no localized calibration" will be removed from the Conclusions. The requirement for annual satellite-specific open-water reference centring and its limitations are now stated explicitly in Sect. 3.3.1 and the Conclusions.

Smaller Issues

Smaller issue 1

> **Smaller issue 1:** *The dataset is cited with two different Zenodo DOIs: 10.5281/zenodo.18452514 in Section 6 (and on the preprint landing page) versus 10.5281/ZENODO.18212535 in the reference list (Song et al., 2026). Please reconcile to a single archived version.*

Response. Thank you for identifying this inconsistency. The DOI ending in 18212535 referred to an earlier archive entry and was retained inadvertently. We have standardized the dataset citation to the archived version used for the revised manuscript: https://doi.org/10.5281/zenodo.18452514. Both the displayed DOI and its underlying hyperlink have been corrected, and the older DOI will be removed.

Changes to be incorporated in the revised manuscript: The dataset DOI has been standardized in Sect. 6 (Data and code availability) and in the Song et al. (2026) reference entry. Both locations now point to https://doi.org/10.5281/zenodo.18452514.

Smaller issue 2

> **Smaller issue 2:** *Typos: double full stop in "polar night (Li et al., 2023).."; stray space in "records remain virtually absent ,"; "1.85°C total temperature increases" should read "increase". Missing space in line 174 – please read through for grammar issues.*

Response. Thank you. We corrected all the listed errors: the duplicated full stop and stray space before the comma were removed, "temperature increases" was changed to "temperature increase," and the missing space was restored. We also performed a manuscript-wide proofreading pass to correct additional punctuation, spacing, subject–verb agreement, and singular–plural inconsistencies.

Changes to be incorporated in the revised manuscript: The identified errors and additional language and formatting inconsistencies have been corrected throughout the Abstract, main text, figure captions, tables, and reference list.

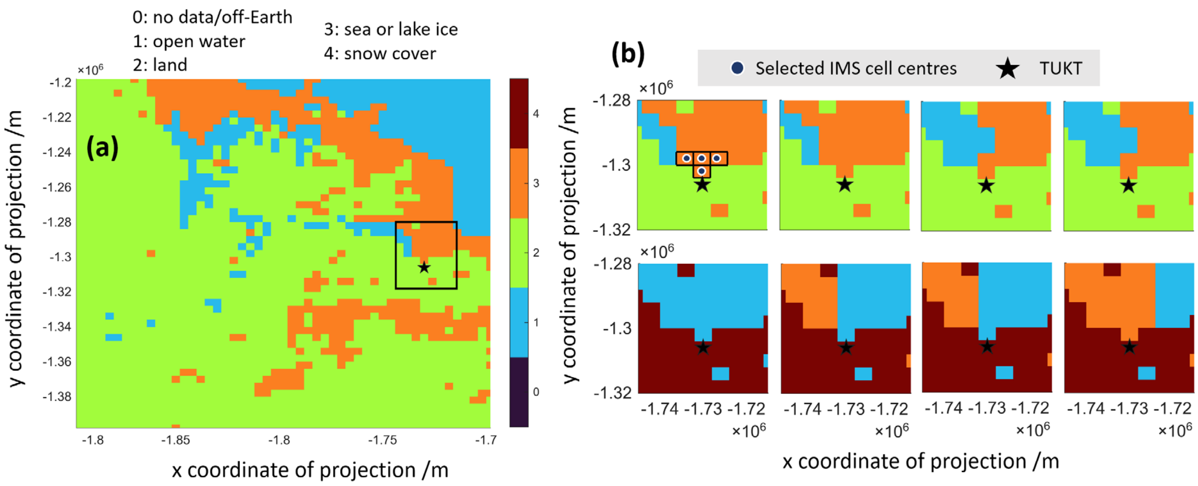
Smaller issue 3

Smaller issue 3: *The IMS window is described as an 11×11 pixel (~44×44 km) region but the comparison is framed against a "4-km product" / "16 km² pixel" elsewhere; a sentence clarifying that the offset statistics use the single central pixel vs the window would avoid confusion.*

Response. We agree that the original description was inconsistent. After rechecking the extraction code and intermediate data, we confirmed that the comparison used neither a single central pixel nor an 11 × 11-pixel window. It used four fixed neighbouring 4 km IMS grid cells surrounding TUKT.

For each day, IMS ice fraction was calculated as the proportion of these four cells classified as ice. The neighbourhood was classified as ice covered when at least two of the four cells indicated ice. The four cells have a nominal combined area of approximately 64 km² within an approximately 12 km × 8 km envelope in the native IMS grid.

The spatial definition is illustrated in Response Figure R1-3. The regional panel shows the TUKT station, the local map extent, and the four selected cell centres, whereas the enlarged panels show their setting during the 2019 ice-out and freeze-up periods. The selected cell locations were fixed across all dates and years.



Response Figure R1-3 Spatial definition of the fixed four-cell IMS neighbourhood and illustrative daily surface classifications near TUKT. (a) Regional 4 km IMS classification map for DOY 157 in 2019. The black star marks the TUKT GNSS station, the black rectangle indicates the local area enlarged in panel (b), and the four blue circles mark the centres of the fixed IMS cells used in the comparison. (b) Enlarged daily IMS classifications for DOY 157–160 (top row) and DOY 292–295 (bottom row) in 2019. The four selected cell centres are shown only in the first local panel to avoid visual clutter. IMS classes are 0, no data/off-Earth; 1, open water; 2, land; 3, sea or lake ice; and 4, snow cover. The same four cells were used for all dates and years, and the local neighbourhood was classified as ice covered when at least two of the four cells were classified as ice.

The references to a single pixel and an 11 × 11-pixel window were reporting errors rather than differences in the implemented analysis. Correcting the description therefore did not alter the IMS time series, detected dates, or comparison statistics.

Changes to be incorporated in the revised manuscript: The four-cell extraction, daily ice-fraction calculation, two-of-four classification rule, and spatial support are now described consistently in Sect. 2.3.2. Corresponding terminology has been standardized in Sect. 4.4, Table 5, Figure 12, and the Conclusions.

Smaller issue 4

Smaller issue 4: *Table 1 lists a period for IMS/ERA5/NAVCAN but leaves Sentinel-2's period blank; add the years/scenes used.*

Response. Thank you. Table 1 will identify 2019 as the optical-satellite comparison period and will state that five selected cloud-free scenes were used: three Sentinel-2A MSI Level-2A scenes and two Landsat 8 scenes. The Sentinel-2A scenes were acquired on 8 June (DOY 159), 16 October (DOY 289), and 20 October (DOY 293), whereas the Landsat 8 scenes were acquired on 9 June (DOY 160) and 15 October (DOY 288), all in 2019.

These scenes are used only to provide qualitative spatial context for the selected 2019 ice-out and freeze-up events. They are not treated as validation of the complete record or as evidence identifying a particular ice type or formation process.

Changes to be incorporated in the revised manuscript: The optical-satellite imagery entry in Table 1 will report "2019; five selected cloud-free scenes (Sentinel-2A, n = 3; Landsat 8, n = 2)." The sensor-specific acquisition dates and qualitative purpose will be specified in Sect. 2.3.4 and the revised Figure 7 caption.

Smaller issue 5

Smaller issue 5: *Figure 1 – Remove the legend with unused symbols, I cannot see the pink square utilised in the figure.*

Response. We agree that the unused legend symbols and repeated legend boxes made the figure unnecessarily cluttered. The redundant legends will be removed. The TUKT station remains identified directly by the magenta marker and adjacent label, and its meaning is stated in the caption. The revised Figure 1 is reproduced below.

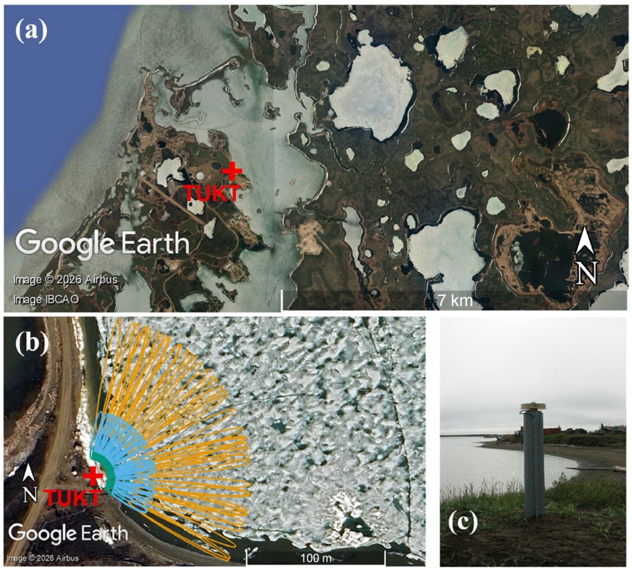


Figure 1. (a) Location of the TUKT GNSS station on the Beaufort Sea coast near Tuktoyaktuk, Northwest Territories, Canada. The magenta marker indicates the station location. (b) Modelled GPS L1 first Fresnel-zone ellipses within

the selected azimuth range of 8°–125° at elevation angles of 5° (orange), 10° (blue), and 15° (green), calculated using an antenna height of 6 m above a flat horizontal reflecting surface. (c) Photograph of the TUKT antenna and monument. Background imagery in panels (a) and (b) is from Google Earth, and the station photograph is from SONEL.

Changes to be incorporated in the revised manuscript: Figure 1 will be revised to remove the unused and redundant legend elements. The TUKT station marker, direct label, and corresponding caption description have been retained.

Smaller issue 6

Smaller issue 6: *Analysis in caption of Figure 3. Analysis should be kept for the text, and captions remain as information as to what is contained in the figure. This is also a problem in figure 6.*

Response. We agree. Interpretive statements will be removed from the figure captions and transferred to the corresponding Methods or Results text.

The revised Figure 3 caption identifies only the antenna-response curves, surface-roughness attenuation, co-polar RHCP Fresnel terms, assumed dielectric properties, and operational elevation interval. The explanation of why the modelled sea-ice response exceeds the seawater response over part of this interval is now confined to Sect. 3.1.2.

The original Figure 6, now Figure 7, has been treated similarly. Its caption identifies the Sentinel-2A and Landsat 8 scenes, acquisition dates, station marker, and regional-overview panel. Interpretation of the spatial development of the 2019 ice conditions is provided in Sect. 4.1 rather than in the caption.

Changes to be incorporated in the revised manuscript: The captions of Figures 3 and 7 have been shortened to identify only the plotted quantities or displayed imagery. Interpretive material has been moved to Sects. 3.1.2 and 4.1. The revised Figure 3 is reproduced in our response to Specific Comment 5, and the revised Figure 7 is shown in our response to Smaller Issue 8.

Smaller issue 7

Smaller issue 7: *Repeated paragraphs – eg. Line 212–215 repeated from line 114 onward; line 412–415 results repeated from above.*

Response. Thank you for identifying these repetitions. The repeated methodological description and duplicated presentation of the results will be removed or consolidated so that each point is introduced once in the most appropriate section. We also checked the remainder of the manuscript for similar repetition and shortened overlapping descriptions where possible.

Changes to be incorporated in the revised manuscript: The repeated material corresponding to the original lines 212–215 and 412–415 has been deleted or merged with its earlier occurrence. The surrounding paragraphs and transitions will be revised to preserve continuity.

Smaller issue 8

> **Smaller issue 8:** *Figure 6 is very hard to follow – put images in 3×2 with DOY marked in the top corner in white; additionally a coastline would help.*

Response. We agree that the original layout was difficult to follow. The figure, now Figure 7, has been redesigned as a 3 × 2 arrangement of six panels derived from five optical scenes: three Sentinel-2A scenes and two Landsat 8 scenes. Each panel is identified consistently by panel letter, acquisition date, and day of year, and the TUKT station is marked throughout.

We also added a regional overview of the 9 June scene, with a rectangle indicating the extent of the corresponding local-scale panel. After testing overlay labels, we retained the date and DOY immediately below each panel because white text was not consistently legible over both bright ice and dark water. A separate coastline vector was not added because the land–water boundary is visible in the enlarged imagery and an additional line obscured nearshore surface features. The regional-overview panel provides the requested geographic context without covering the imagery.

The redesigned Figure 7 is reproduced below.

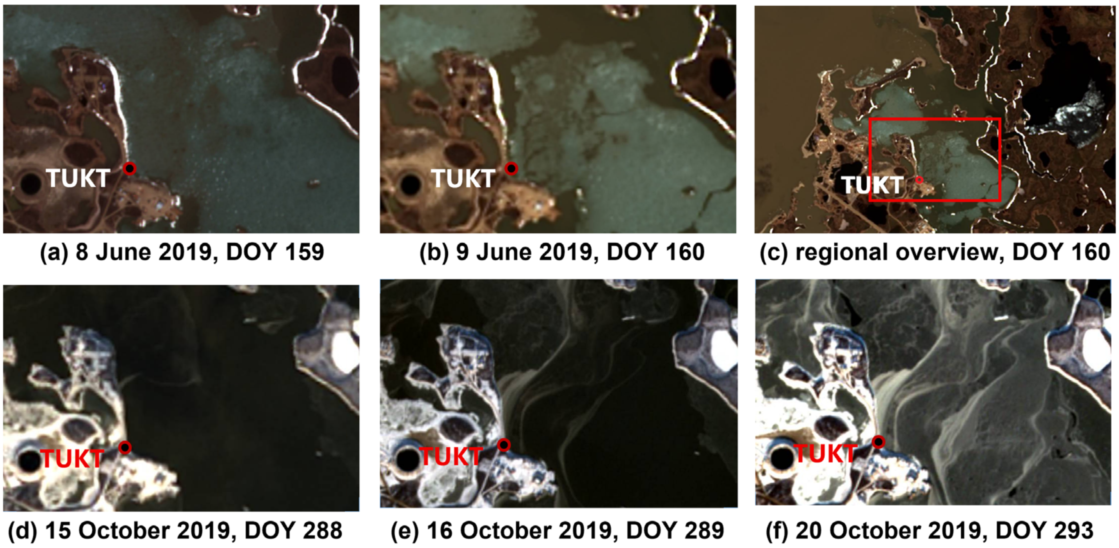


Figure 7. Optical satellite imagery around the TUKT station during the spring and autumn sea-ice transition periods in 2019. (a) Sentinel-2A image on 8 June (DOY 159); (b) Landsat 8 image on 9 June (DOY 160); (c) regional Landsat 8 image on 9 June (DOY 160), with the red rectangle indicating the local area shown in panel (b); (d) Landsat 8 image on 15 October (DOY 288); (e) Sentinel-2A image on 16 October (DOY 289); and (f) Sentinel-2A image on 20 October (DOY 293). The red circle indicates the TUKT GNSS station. Data sources: Copernicus Sentinel-2A and Landsat 8.

Changes to be incorporated in the revised manuscript: The original Figure 6 has been replaced by the 3 × 2 Figure 7, with enlarged panels, consistent station markers, date and DOY labels, and a regional-overview panel linked to the local view by a marked footprint.

Smaller issue 9

Smaller issue 9: *Figure 7 – grey shading hard to pick out on top of the purple. This is the key figure, the time series should extend the width of the page, the purple points maybe need to be smaller in order that they are more easily visible, and the whole thing a lot bigger.*

Response. We agree that the complete AIF time series is a central figure and required greater visual prominence. The original Figure 7, now Figure 8, has been divided into two wide panels covering 2003–2012 and 2013–2023. Both panels extend across the available page width, avoiding compression of two decades of sub-daily observations into a single narrow axis.

The individual satellite-track observations are now plotted using smaller, lighter purple points. The filtered AIF is shown with a distinct black line, and the gray residual-variability envelope has been made more visible relative to the observations. The revised figure is reproduced below.

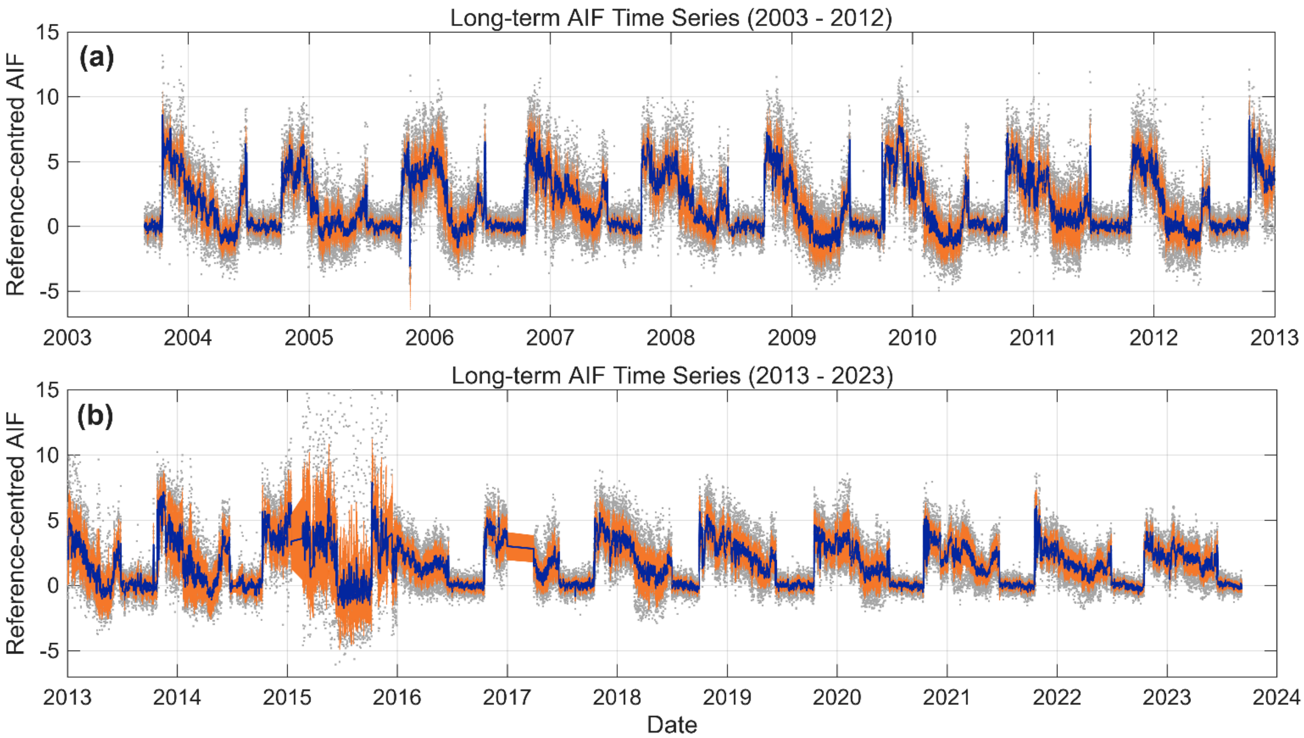


Figure 8. Complete reference-centred AIF record at TUKT for (a) 2003–2012 and (b) 2013–2023. Light-purple points show individual satellite-track observations, the black line shows the first-order Butterworth low-pass-filtered AIF, and gray shading denotes ±1 moving residual standard deviation calculated over 24 h.

Changes to be incorporated in the revised manuscript: Figure 8 has been enlarged and reformatted as two full-width time-series panels. The observation-point size and transparency, filtered-line contrast, and visibility of the gray residual-variability envelope have been adjusted.

Smaller issue 10

Smaller issue 10: *Figure 7 caption "a filtered AIF" – what filter? A rolling mean or something else?*

Response. Thank you. The filtering method is now identified explicitly. The black curve in the revised Figure 8 is the reference-centred AIF processed using a first-order Butterworth low-pass filter with a cutoff frequency of 1.0

cycle d^{-1}.

The gray envelope is not a confidence interval around the filtered mean. It represents ±1 moving standard deviation, calculated over 24 h, of the residuals between the individual reference-centred AIF observations and the filtered curve.

Changes to be incorporated in the revised manuscript: The filtering method, filter order, cutoff frequency, and definition of the gray residual-variability envelope are now stated in Sect. 3.3.1 and the Figure 8 caption. The revised Figure 8 is reproduced in our response to Smaller Issue 9.

Smaller issue 11

Smaller issue 11: *Figure 8 – orange border looks a lot like the red negative anomaly bars – find a more unique colour to avoid confusion.*

Response. We agree that the orange border could be mistaken for a data category and was not necessary for interpreting the figure. Rather than replacing it with another prominent colour, we removed the border entirely. Panel separation is now provided by the subplot layout and neutral axes, while the plotted quantities are distinguished by their marker shapes and line colours.

Changes to be incorporated in the revised manuscript: The orange border will be removed from the revised long-term tendency figure, now Figure 9. The revised figure is reproduced in our response to Major Comment 1.

Smaller issue 12

Smaller issue 12: *Line 387 "Crucially…cycle" is repetition from 354.*

Response. Thank you. The sentence beginning with "Crucially" repeated a conclusion already stated earlier and will be removed. The surrounding paragraph has been consolidated so that it proceeds directly from the numerical trend estimates to their statistical interpretation without restating the same conclusion.

Changes to be incorporated in the revised manuscript: The repeated sentence has been deleted from Sect. 4.2.2, and the surrounding paragraph will be revised for continuity and concision.

Smaller issue 13

Smaller issue 13: *Line 390 "decade) is a genuine, persistent climatological shift rather than a statistical artifact driven by outlier years." – Large claim please evidence, see Point 1 above.*

Response. We agree that this statement was not supported by the record length or statistical uncertainty and have removed it. We also no longer exclude 2018 or describe it as an outlier. All years with valid observations are retained in the primary trend analysis.

Under the primary stable–stable definition, corrected cross-year ice-season duration tended to decrease at −3.31 d $decade^{-1}$, with a 95 % confidence interval from −7.82 to +0.63 d $decade^{-1}$ and $p = 0.086$. Because the confidence interval includes zero and the Mann–Kendall test does not reach the 0.05 significance level, the result is now described as a tendency toward a shorter continuous ice season rather than a demonstrated persistent climatological shift.

One alternative boundary definition produced a borderline estimate of −4.23 d decade^{-1} ($p = 0.048$), but this result was not reproduced by the other three boundary definitions and is therefore presented only as a sensitivity result. The complete statistical reframing is described in our response to Major Comment 1.

Changes to be incorporated in the revised manuscript: The unsupported statement will be removed from Sect. 4.2.2. The Abstract, Sect. 4.2.2, Table 4, Figure 9, and Conclusions will distinguish directional tendencies from statistically significant trends and report the corresponding uncertainty.

Smaller issue 14

Smaller issue 14: *Several conclusions are stated causally without support. Please evidence or soften the statements concerning thermal memory, validation of temporal accuracy, spatial heterogeneity, and physical detection thresholds.*

Response. We agree and have revised each of these statements. The attribution to "thermal memory" will be removed because the available data do not isolate that mechanism from other thermal and dynamic controls. Freeze-up and ice-out timing are now discussed as being associated with the timing of thermal forcing while potentially also reflecting wind-driven ice movement, storms, snow conditions, fracturing, and other coastal processes.

The terms "validates" and "temporal accuracy" have also been removed where the analyses provide physical correspondence rather than independent validation. The degree-day relationships are now described as supporting the physical consistency of the phenological record but not as independently validating the individual transition dates.

The GNSS-IR–IMS timing differences are no longer attributed to "extreme spatial heterogeneity." They are described as potentially reflecting a combination of different spatial supports, sensitivity to partial or newly formed ice cover, and actual nearshore spatial variability. The manuscript states explicitly that their relative contributions cannot be quantified from the available observations.

Finally, the claim that the timing differences represent different physical detection thresholds rather than measurement errors will be removed. The revised discussion will state only that the differences are consistent with the distinct spatial supports, temporal resolutions, and observational definitions of GNSS-IR, IMS, and the optical imagery, while ERA5 provides meteorological context rather than an independent sea-ice state measurement.

Changes to be incorporated in the revised manuscript: Unsupported causal statements will be revised in Sects. 4.1, 4.3, and 4.4 and in the Conclusions. Definitive terms such as "validates," "thermal memory," and "represent" have been replaced with qualified formulations such as "is associated with," "may reflect," "is consistent with," and "supports."

Smaller issue 15

Smaller issue 15: *LaTeX has leaked into the text: "strict statistical significance ($p < 0.05$)" in Section 4.2.2.*

Response. Thank you. The residual LaTeX notation and the unnecessarily emphatic phrase "strict statistical significance" will be removed. The sentence is now expressed in plain text as: "None of the configuration-specific Mann–Kendall tests had a p value below 0.05." We also checked the manuscript for other mathematical markup inadvertently retained in normal prose.

Changes to be incorporated in the revised manuscript: The affected sentence in Sect. 4.2.2 has been corrected, and residual LaTeX delimiters and escaped mathematical symbols will be removed from the normal text throughout the manuscript.

Smaller issue 16

Smaller issue 16: *ERA5 vs in-situ agreement is reported both as "R = 99.1%" and "r = 0.99". Report a correlation coefficient (r = 0.99), not a percentage, and use it consistently.*

Response. We agree. This quantity is a Pearson correlation coefficient rather than a percentage. We removed "R = 99.1%" and will report the agreement between daily ERA5 and temporally matched Tuktoyaktuk Airport air temperatures consistently as r = 0.99, together with an RMSE of 1.98 °C.

Changes to be incorporated in the revised manuscript: Sects. 3.3.4 and 4.3.1 will be revised to use Pearson correlation notation consistently and to report the corresponding RMSE.

Smaller issue 17

Smaller issue 17: *Figure 9 caption should be made more concise to aid legibility of figures. Long sentence descriptions delay comprehension of figures.*

Response. We agree. The caption has been shortened to identify only the plotted variables, the common 2004–2022 analysis period, and the trend estimators. Interpretation of the interannual scatter and the limitations of annual mean temperature has been retained in the main text rather than repeated in the caption.

Changes to be incorporated in the revised manuscript: The original Figure 9 is now Figure 10. Its caption has been shortened, and the interpretive discussion is provided in Sect. 4.3.1.

Smaller issue 18

Smaller issue 18: *Figure 10 caption unclear. Subplots have headings below and repeated above but only for some of them. Stated as Heatmaps when they are scatter plots (or line charts). White stars referenced in caption but only coloured stars present in figures.*

Response. We agree and revised both the panel labelling and caption. The original Figure 10 is now Figure 11. Panels (a) and (b) are identified as line plots of the Pearson correlation coefficient across accumulation-window positions and durations, whereas panels (c) and (d) are identified as scatter plots for the selected TDD and FDD windows. The panel headings are now consistent, and the caption refers simply to stars marking the selected windows without assigning an incorrect colour.

The revised Figure 11 is presented with our response to Major Comment 4, where the associated window-search analysis is discussed.

Changes to be incorporated in the revised manuscript: Figure 11 and its caption will be revised to distinguish the two plot types, standardize the panel headings, and correct the description of the selected-window markers.

Smaller issue 19

Smaller issue 19: *Table 6 – seems unnecessary to list all 4 results of this study in the review summary table.*

Response. We agree. Listing several variations of the TUKT estimate repeated results already reported in Table 4 and made the benchmarking table unnecessarily long. Table 6 has therefore been reduced to two rows from the present analysis: the primary full-record GNSS-IR corrected ice-season trend and one IMS contextual-comparison row. The common-period GNSS-IR and IMS estimates are now summarized in the accompanying text rather than repeated as separate table entries.

The revised Table 6 is presented with our response to Specific Comment 14. It is used only to place the local TUKT tendency in the context of published Arctic changes; the text explicitly states that differences in period, spatial scale, sensing method, and metric definition preclude direct equivalence or independent validation.

Changes to be incorporated in the revised manuscript: Table 6 has been shortened to two rows from the present analysis, and the common-period estimates have been moved to the accompanying discussion in Sect. 4.5.

We thank the reviewer again for the careful and constructive comments. Our final responses document the additional analyses completed in response to the review, including complete-record non-parametric trend estimation, symmetric and gap-aware transition detection, detector-configuration and threshold-sensitivity tests, receiver-period diagnostics, the AIF–AMP comparison, window-search-adjusted degree-day tests, and scale-aware comparison with IMS and optical imagery. They also distinguish direct evidence from physically plausible interpretation and retain all valid years, including 2018, in the primary analyses. If the editor invites a revised manuscript, we will incorporate the corresponding textual, methodological, figure, and table changes at the locations identified in each response. These changes will improve the statistical transparency, methodological consistency, and reproducibility of the study.