Review #2

Summary

This manuscript uses the Amplitude Integration Factor (AIF) to analyze a 20-year GNSS-IR record of coastal sea ice phenology at the TUKT station and examines long-term changes in ice-out, freeze-up, and ice-season duration. The long observation period, the focus on the poorly observed nearshore environment, and the attempt to homogenize GNSS observations across receiver changes are potentially valuable. The comparison with IMS, ERA5, in situ meteorological observations, and Sentinel-2 imagery also provides useful multi-scale context.

I have also reviewed the authors' previous work describing the development and validation of the AIF indicator. Accordingly, the methodological novelty of AIF itself is not a central issue in my evaluation of this manuscript. The key question here is whether this indicator can be reliably extended to a 20-year climate-scale record, and whether the resulting phenological dates and long-term trends are supported by sufficiently rigorous calibration, uncertainty assessment, and independent validation.

At present, several central conclusions appear stronger than the analyses can support. In particular, the long-term climatological conclusion depends strongly on the treatment of an individual year, while the hardware homogenization, transition-date detection, and physical interpretation of AIF require stronger validation. I therefore recommend major revision before the main conclusions can be considered sufficiently robust.

General response

We thank the reviewer for the careful and constructive assessment and for focusing the evaluation on whether AIF can support a consistent two-decade phenological record. We agree that the original submission did not sufficiently separate the previously developed AIF indicator from the contribution of the present study and that several interpretations were stronger than the available evidence.

We have verified the first Fresnel-zone geometry and four-cell IMS extraction, completed a direct observed elevation-profile analysis, documented the equipment history, quantified receiver-period AIF differences, tested transition-date homogeneity, reprocessed the record using equivalent direction-reversed transition criteria, and completed a like-for-like comparison between AIF and conventional Lomb–Scargle peak amplitude.

Our reanalysis retains all valid years, including 2018. The primary tendencies are estimated using Theil–Sen slopes and two-sided Mann–Kendall tests; none reaches the 0.05 significance level. All ice-season-duration results use the corrected cross-year interval from freeze-up in year $y - 1$ to ice-out in year y.

The 2019 optical-satellite sequence, comprising three Sentinel-2A scenes and two Landsat 8 scenes, is treated as qualitative spatial context. Thermodynamic results are described as associations rather than proof of ocean heat storage, and the GNSS-IR–IMS difference is interpreted as a cross-scale discrepancy without a uniquely identified cause. If the editor invites a revised manuscript, we will incorporate the corresponding changes to the title, Abstract, Methods, Results and discussion, Conclusions, figures, tables, and captions as specified below.

Major Comments

Major comment 1: Section 2.2: GNSS-IR reflection footprint

Status: Addressed

Reviewer comment. *Section 2.2 states that the reflection footprint ranges from approximately 8 to 40 m in radius and covers 50–400 m2. These values appear geometrically inconsistent. Moreover, the First Fresnel Zone is elliptical rather than circular, so describing it by a single radius may be misleading. The authors should verify these values and provide a consistent description of the spatial footprint. If the exact dimensions are not essential to the subsequent analysis, I suggest removing the specific radius and area values rather than introducing potentially confusing spatial-scale estimates.*

Response. We agree and recalculated the GPS L1 first Fresnel-zone geometry using a wavelength of approximately 0.190 m, an antenna height of 6 m, a flat horizontal surface, and the final 5°–15° elevation interval. Each footprint is now described as an ellipse rather than by a single radius.

At 5° elevation, the semi-major and semi-minor axes are 43.38 and 3.78 m, the ellipse centre is 81.06 m from the antenna along the satellite azimuth, and the area is 515.17 m². At 15°, the corresponding values are 8.24 m, 2.13 m, 23.76 m, and 55.19 m². These are dimensions of individual Fresnel-zone ellipses, not of the complete multi-azimuth sampling sector.

We also rechecked the ellipses against the coastline. The revised manuscript states that the selected sector primarily samples the nearshore surface while acknowledging that limited contributions from the immediate coastal margin cannot be completely excluded.

The verified end-member dimensions are summarized below.

Elevation	Semi-major axis	Semi-minor axis	Centre distance	Ellipse area
5°	43.38 m	3.78 m	81.06 m	515.17 m²
15°	8.24 m	2.13 m	23.76 m	55.19 m²

The revised Figure 1 below shows the elliptical footprints and their coastal setting.

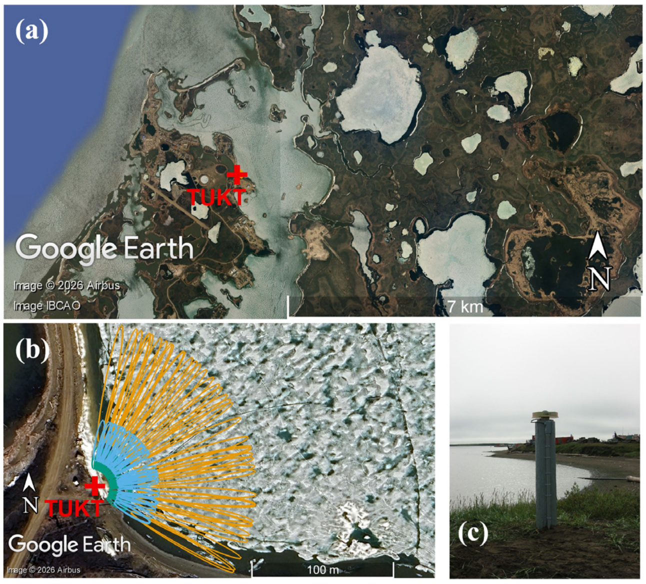


Figure 1. (a) TUKT station on the Beaufort Sea coast near Tuktoyaktuk, Canada. (b) Modelled GPS L1 first Fresnel-zone ellipses for 8°–125° azimuth at 5° (orange), 10° (blue), and 15° (green) elevation, assuming a 6 m antenna height and flat surface. (c) TUKT antenna and monument. Background imagery: Google Earth; station photograph: SONEL.

Changes to be incorporated in the revised manuscript: The radius-based description in Sect. 2.2 will be replaced by verified semi-axis, centre-distance, and area ranges. The calculation assumptions and coastal-margin limitation will be added, and Figure 1 and its caption will be revised.

Major comment 2: Section 2.3.2: Definition of the IMS comparison

Status: Addressed

Reviewer comment. *The IMS processing requires further clarification. Section 2.3.2 defines regional ice presence using an 11×11 pixel window, corresponding to approximately 44×44 km, with an "ice probability" threshold of 0.5. However, later sections discuss the comparison in terms of a single 4-km IMS grid cell or a 16-km2 pixel. These represent substantially different spatial supports.*

The authors should clearly specify whether the phenological dates are derived from the central IMS pixel, an 11×11 pixel window, or a spatially aggregated ice fraction, and explain how the 0.5 threshold is calculated. This distinction is essential for interpreting the subsequent comparison with the much smaller GNSS-IR reflection footprint.

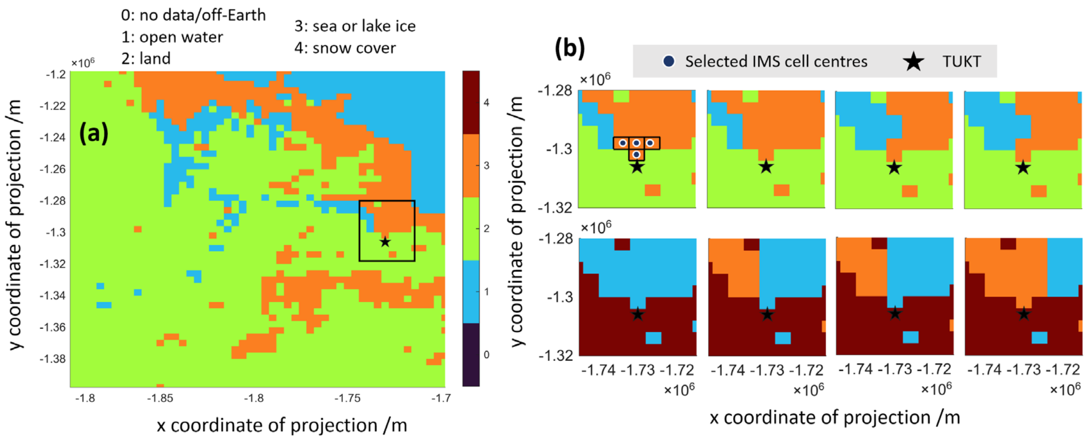
Response. We agree that the original descriptions were inconsistent. Rechecking the extraction code, intermediate data, and plotting records confirmed that the comparison used neither a single central cell nor an 11 × 11-cell window. It used four fixed neighbouring 4 km IMS grid cells adjacent to the TUKT nearshore sector. The same four cells were used for every daily map and throughout the IMS comparison period.

For each day t, the local IMS ice fraction was calculated as follows:

$$F_{\mathrm{IMS}}(t) = \frac{N_{\mathrm{ice}}(t)}{4}$$

where $N_{\mathrm{ice}}(t)$ is the number of the four selected cells classified as sea or lake ice (IMS class 3). The resulting fraction therefore ranges from 0 to 1 in increments of 0.25. An ice-covered state was assigned when $F_{\mathrm{IMS}} \geq 0.5$, corresponding to at least two of the four cells being classified as ice. The four cells have a nominal combined area of approximately 64 km² and occupy an approximately 12 km × 8 km envelope in the native IMS grid.

To make the extraction geometry explicit, Response Figure R2-1 shows both the regional setting and the local daily classifications. The regional panel identifies TUKT, the local map extent, and the centres of the four selected IMS cells. The enlarged panels illustrate the evolution of the IMS classifications during the 2019 ice-out and freeze-up periods. The four blue markers in the first local panel are the fixed grid-cell centres used in the calculation; they are shown only once to avoid obscuring the daily classifications.



Response Figure R2-1. 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 earlier references to a single cell and an 11 × 11-cell window were reporting errors rather than alternative processing methods. Correcting these descriptions did not alter the underlying IMS series, detected dates, or comparison statistics. IMS is now consistently presented as a neighbourhood-scale comparison product rather than as point-scale ground truth, and F_{IMS} is described as a local neighbourhood ice fraction rather than an ice probability.

Changes to be incorporated in the revised manuscript: The exact four-cell extraction, daily ice-fraction calculation, two-of-four threshold, nominal spatial support, and scale-aware interpretation will be added to Sect. 2.3.2 and applied consistently in Sect. 4.4, Table 5, Figure 12, and the Conclusions.

Major comment 3: Sections 3.1–3.2: Role of AIF and observational validation of its physical basis

Status: Addressed

Reviewer comment. *AIF was previously developed and validated by the authors in Song et al. (DOI: 10.1109/TGRS.2022.3155051). Although AIF remains central to the present analysis, an extensive re-derivation of the indicator does not appear necessary here. Citing the previous study and briefly describing the principle and implementation used in the present work should be sufficient. If the AIF calculation differs from the previous study, these differences should be clearly identified and justified. I therefore do not regard the development of AIF itself as a primary methodological contribution of the present manuscript.*

Instead, the manuscript would benefit more from showing the observed elevation-dependent amplitude curves, P (Eq. 4), under different sea-surface states. In particular, do the observed dSNR amplitudes show the same elevation-dependent behavior as the theoretical or simulated results in Figure 3? For example, is the observed amplitude over sea ice consistently higher than that over open water throughout the 5°–15° elevation range? Such an observational comparison would provide more direct support for the physical basis of the AIF analysis than simply presenting integrated AIF values for ice and water.

Response. We agree. The Introduction now attributes the development and validation of AIF to Song et al. (2022), and the method section retains only the definition and implementation needed for this study. The AIF definition itself is not claimed as new; the present study applies the established formulation, while the long-record quality control, annual reference centring, and gap-aware phenology detection are identified as processing steps for this analysis. The contribution here is the consistently processed two-decade record and its transition, sensitivity, homogeneity, and cross-scale analyses.

We added a direct observed-profile test by matching stable-ice and open-water tracks within year, PRN, and azimuth cluster, with at least five tracks per state. After excluding incomplete endpoint years and mixed-receiver 2015, the analysis contained 18 complete years and 637 matched groups, with one equally weighted profile pair per year.

The median ice-minus-water difference was positive over 80.2 % of the elevation range, and its pointwise 95 % lower limit was positive over 74.4 %. The integrated contrast was positive in all 18 years (two-sided sign-flip $p < 0.001$) but weakened above approximately 13°. We therefore support integration across 5°–15° without claiming a higher ice response at every elevation angle.

The empirical comparison is shown in the revised Figure 4.

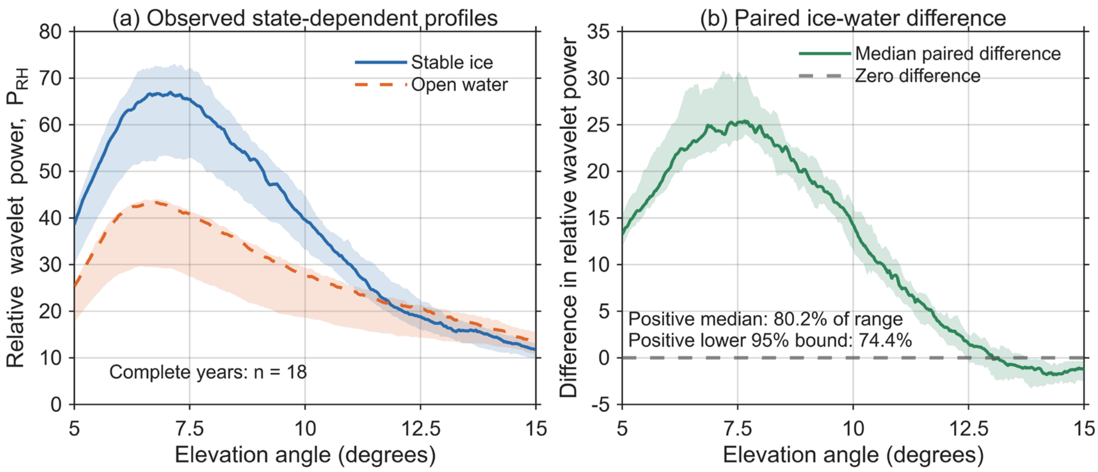


Figure 4. Observed elevation-dependent relative wavelet-power profiles for stable ice and open water. (a) Median profiles across 18 complete and equally weighted observation years from 2004 to 2022, excluding mixed-receiver 2015. (b) Median paired ice-minus-water difference. Shading denotes pointwise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

Changes to be incorporated in the revised manuscript: The novelty framing has been shortened in the Introduction and Sect. 3.2. The observed-profile matching and statistical test will be added to Sects. 3.2 and 4.1, and Figure 4 and its caption now show the empirical stable-ice and open-water profiles.

Major comment 4: Section 3.3.1: Inter-calibration across receiver changes

Status: Addressed

Reviewer comment. *The manuscript reports an approximately 27% reduction in AIF magnitude following the 2015 receiver change and subsequently applies satellite-specific subtraction of the mean AIF during DOY 200–250. However, this normalization removes an additive offset, whereas the hardware effect is described as a proportional reduction in signal magnitude.*

The authors should clarify whether the receiver-induced change is additive, multiplicative, or more complex, and demonstrate that the adopted normalization adequately removes its effect on both the open-water baseline and the seasonal ice–water contrast. In addition, the manuscript states that several hardware changes occurred during the 20-year period, while only the 2015 change is discussed in detail. A complete equipment history and evidence that other hardware changes do not introduce detectable discontinuities would strengthen the reliability of the long-term record.

Response. We agree that the original description conflated an observed period difference with a proportional receiver model. The implemented normalization is additive: for each satellite and year, the mean AIF over the DOY 200–250 open-water reference interval is subtracted from that year's observations. This aligns the annual open-water reference level but does not assume that winter AIF and the summer baseline are scaled proportionally.

Using annual November–December means of the reference-centred, filtered AIF, the pre-replacement mean was 4.36 ± 0.47 for 2003–2014 and the post-replacement mean was 3.25 ± 0.50 for 2016–2022, a 25.6 % reduction. The corresponding source-timestamp and daily-median checks gave 25.7 % and 25.1 %, respectively, showing that the difference was not introduced by fixed-grid interpolation. Because the receiver periods contain different years and environmental conditions, the full amplitude difference is not attributed exclusively to the receiver change.

The site log confirms that the ASHTECH UZ-12 was replaced by a TRIMBLE NETR9 on 16 December 2015; other listed changes during the analysis period were firmware updates, while the antenna, radome-free configuration, cable, monument, and marker eccentricities remained unchanged. A SEPT POLARX5 was installed only after the last usable AIF observation in 2023.

Direct homogeneity tests found no detectable transition-date step. Detrended Mann–Whitney p values were 0.728 for ice-out and 0.874 for freeze-up. Segmented-regression step estimates were +1.98 d for ice-out (95 % CI −6.50 to +10.47 d; $p = 0.628$) and −2.57 d for freeze-up (95 % CI −13.75 to +8.61 d; $p = 0.634$). These tests do not establish perfect amplitude homogeneity, and the limitation is stated explicitly.

Changes to be incorporated in the revised manuscript: The receiver and firmware history has been clarified in Sect. 2.2. Sect. 3.3.1 now distinguishes additive reference centring from a multiplicative hardware model and reports the 25.6 % period difference, source-data sensitivity checks, and transition-date homogeneity tests; the limitation is carried into Sect. 4.2.1 and the Conclusions.

Major comment 5: Section 3.3.2: Phenological transition criteria and the added value of AIF

Status: Addressed

Reviewer comment. *The transition-date criteria require further clarification. With σ_ref=0.19, the nominal detection threshold is 3σ_ref=0.57 in calibrated AIF units. However, ice-out additionally requires a "sustained low value," for which no duration is specified, whereas freeze-up requires AIF to remain above the threshold for at least five days. These asymmetric persistence criteria introduce some subjectivity and should be defined more explicitly.*

They also suggest that the method is primarily intended for retrospective phenological analysis. If so, the authors should clarify the specific advantage of AIF over conventional spectral-amplitude metrics. Although spectral amplitude may provide weaker instantaneous ice–water separation, under the same smoothing, thresholding, and persistence-based post-processing it may potentially provide comparable transition dates. A direct comparison of AIF and conventional spectral amplitude in terms of state separability and transition-date stability would better demonstrate the added value of AIF for long-term phenological analysis.

Response. We agree that the transition rules in the original submission were asymmetric and insufficiently specified. The baseline AIF state threshold was retained at $\eta = 3\sigma_{\mathrm{ref}} = 0.570$, but the complete record was reprocessed using equivalent, direction-reversed criteria for ice-out and freeze-up.

For both stable transitions, at least 60% of the expected observations were required during each of the preceding and following 7 d intervals. At least 50% of the valid observations in the preceding interval had to represent the previous seasonal state, whereas at least 90% of those in the following interval had to represent the new state. Longer-term confirmation required 14 equivalent observed days within no more than 35 calendar days, with at least 90% of the valid observations representing the new state. Equivalent direction-reversed terminal-anchor, 5 d sustained-reversal, gap-handling, and candidate-de-duplication rules were then applied. Only the source and target states and their physically appropriate seasonal search intervals differed. No transition date was assigned manually.

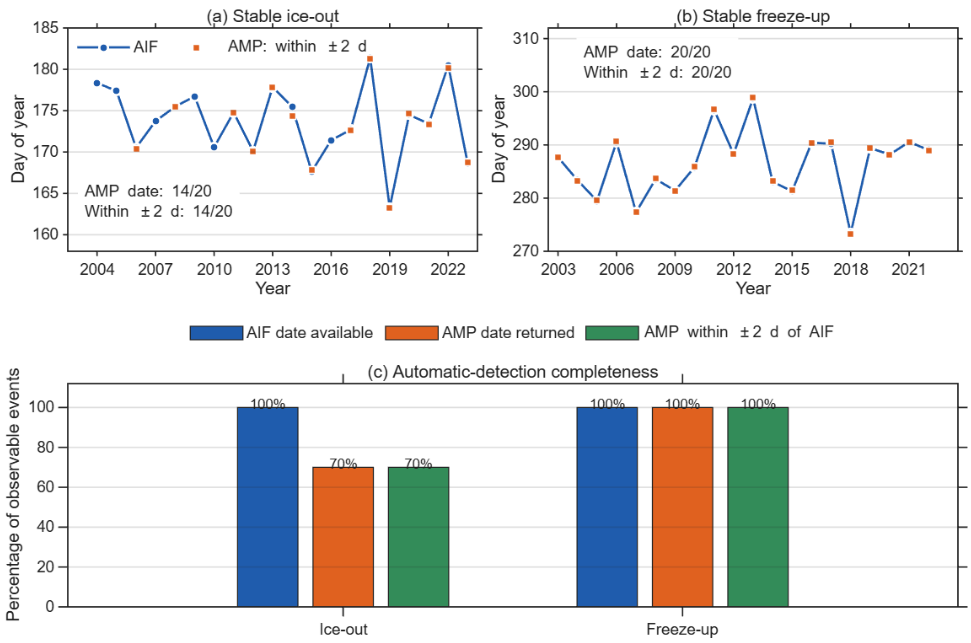
In addition to the stable boundaries used in the primary analysis, early ice-out and freeze-up were extracted to describe the first supported transition attempt. Both early events used a preceding 3 d window with at least 80% coverage and 70% support for the preceding state, followed by a common 2 d persistence interval with at least 80% coverage and 90% support for the new state. When no distinct early event was detected, the early and stable boundaries coincided. Early boundaries were used only to characterize transition development and evaluate boundary-definition sensitivity; the primary phenological results used the stable boundaries.

Dependence on the selected temporal-support parameters was evaluated using all $3 \times 3 \times 2 = 18$prespecified combinations of independently varying stable ice-out and freeze-up persistence among 3, 5, and 7 d and setting longer-term confirmation to either 10 or 14 equivalent observed days. All configurations preserved the directions of the principal trends, and none of the configuration-specific Mann–Kendall tests reached $p < 0.05$.

We also performed the direct AIF–AMP comparison suggested by the reviewer. AIF and conventional dominant Lomb–Scargle spectral-peak amplitude (AMP) were reference-centred using the same open-water calibration window, processed using identical filtering and daily aggregation, and evaluated under the same seasonal search intervals and detector rules. Each indicator used a state threshold scaled by its own open-water standard deviation. Across 6,353 matched daily values, AIF and AMP were strongly correlated ($r = 0.928$).

State separability was quantified as the annual difference between the stable-ice and stable-water medians divided by their pooled robust within-state scatter. Across the 18 complete single-hardware years available for this paired diagnostic comparison, the median state separation was 2.04 for AIF and 1.69 for AMP. A two-sided paired sign-flip permutation test of the annual differences, based on 9,999 permutations, gave $p = 0.029$. The years 2003 and 2023 were excluded from this diagnostic because of incomplete annual coverage, and 2015 was excluded because it spanned the receiver change; these exclusions were not applied to the event-completeness or primary trend analyses.

The date comparison provided a clearer practical distinction. AIF returned all 20 observable stable ice-out dates, whereas AMP returned 14 of 20 (exact McNemar $p = 0.031$). All 14 matched AMP ice-out dates were within 2 d of the corresponding AIF dates; the mean AMP-minus-AIF difference was −0.12 d, the RMSE was 0.33 d, and the maximum absolute difference was 1.15 d. Both indicators returned all 20 observable freeze-up dates, and all were within 2 d; the corresponding mean difference, RMSE, and maximum absolute difference were +0.06, 0.12, and 0.34 d, respectively. The annual comparisons and automatic-detection completeness are shown in Response Figure R2-2.



Response Figure R2-2. Like-for-like comparison of stable transition dates returned by AIF and AMP under identical temporal processing and detector rules. (a) Stable ice-out dates; (b) stable freeze-up dates; and (c) percentages of AIF-observable events for which AIF returned a date, AMP returned a date, and the AMP date was within 2 d of the AIF date.

These results provide a more specific and limited interpretation of the added value of AIF. AMP produced closely comparable dates whenever it returned a transition, especially for freeze-up. The main advantage of AIF was therefore not a systematic displacement of the detected dates, but stronger ice–water state separation and more complete automatic detection of the comparatively gradual and variable ice-out transition. Because AIF and AMP were derived from the same GNSS observations, this comparison is an internal method comparison rather than independent validation.

Changes to be incorporated in the revised manuscript: Section 3.3.2 and Table 2 will specify the equivalent direction-reversed stable and early transition criteria, including temporal coverage, state support, longer-term confirmation, terminal anchors, reversal rejection, gap handling, and candidate de-duplication. The Methods and Results will also include the like-for-like AIF–AMP comparison, including state separation, transition-date completeness, and matched-date differences. The interpretation of AIF's added value will be revised to emphasize improved automatic ice-out completeness rather than universally different transition timing.

Major comment 6: Section 4.1: Physical interpretation of the 2019 seasonal phases

Status: Addressed

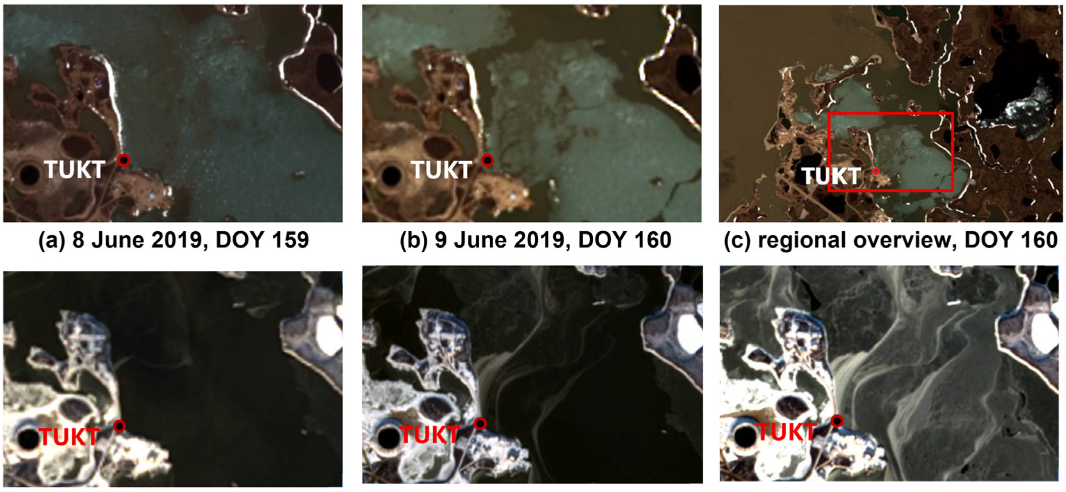
Reviewer comment. *The physical interpretation of the five AIF phases appears more specific than the independent observations can support. Several processes, including snow densification, destructive metamorphism, exposure of bare ice, melt-pond formation, and changes in dielectric loss, are inferred without direct field observations.*

Based on the temperature evolution and satellite imagery presented in the manuscript, Phase II and Phase III appear more straightforwardly interpretable as snow-melt-dominated and ice-melt-dominated periods, respectively. I suggest that the authors distinguish more clearly between surface states directly supported by independent observations and physical processes inferred from the AIF behavior. The phase-specific interpretation should be moderated unless additional independent evidence is available.

Response. We agree. AIF responds jointly to surface state, roughness, snow and ice properties, geometry, and antenna response; the five phases cannot uniquely isolate snow metamorphism, wetness, melt ponds, bare ice, or dielectric loss. In line with the reviewer's suggestion, Phase II and Phase III are now described more conservatively as 'melt onset' and 'advanced melt,' respectively, rather than as uniquely diagnosed process states. The revised discussion separates directly observed changes from possible explanations and uses conditional language throughout.

The 2019 optical-satellite sequence, comprising three Sentinel-2A scenes and two Landsat 8 scenes, is treated as qualitative event-scale context. It shows the spatial progression of surface conditions in five cloud-free scenes but does not identify frazil ice, validate the complete record, or determine the cause of the multi-year GNSS-IR–IMS difference. Statements such as 'perfectly coincides' and 'ultimate validation' have been removed.

The reverified five-scene sequence is reproduced below for clarity.



(a) 8 June 2019, DOY 159 (b) 9 June 2019, DOY 160 (c) regional overview, DOY 160

(d) 15 October 2019, DOY 288 (e) 16 October 2019, DOY 289 (f) 20 October 2019, DOY 293

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 five-phase discussion in Sect. 4.1 and the text around Figures 5–7 have been rewritten using conditional language and explicit observational limitations. The verified Sentinel-2A and Landsat 8 scenes, acquisition dates, and sensor identities will be specified in the revised Figure 7 and caption.

Major comment 7: Section 4.2.2: Treatment of the 2018 anomaly and significance of the long-term trend

Status: Addressed

Reviewer comment. *The exclusion of 2018 currently appears to be motivated primarily by its strong influence on the fitted regression and its characterization as a meteorological anomaly. When the complete record is retained, the corrected ice-duration trend is −4.19 days per decade with p=0.06, whereas the statistically significant value of −4.63 days per decade with p=0.04 is obtained only after excluding 2018.*

An extreme year in a 19-year climate record is not necessarily an invalid observation. It may represent genuine coastal climate variability and should not be removed solely because its inclusion reduces statistical significance. The full-record result should therefore remain the primary estimate, while the result excluding 2018 can be presented as a sensitivity analysis unless an independent criterion demonstrates that the 2018 observation is invalid. The corresponding statements in the Abstract, Results, and Conclusions should be revised accordingly.

Response. We agree. We recalculated all primary analyses using the complete record, retaining 2018 and every other valid year. No observation was excluded because of its effect on the estimated slope or p value. The exclusion-based significant result is no longer used, and the characterization of 2018 as “extreme noise” will be removed from a revised manuscript.

The complete-record primary results were calculated using Theil–Sen slopes and two-sided Mann–Kendall tests. Stable ice-out, stable freeze-up, ice-free period, and corrected cross-year ice-season duration changed by −1.97, +2.85, +5.36, and −3.31 d decade^{-1}, with p values of 0.347, 0.284, 0.294, and 0.086, respectively. All corresponding 95% confidence intervals included zero. Corrected cross-year ice-season duration was calculated from 20 annual values covering 2004–2023 and is therefore interpreted as an uncertain tendency rather than an established shift.

To evaluate sensitivity to the choice of seasonal boundary, we also extracted early ice-out and freeze-up dates. An early boundary represents the first supported transition attempt into the new seasonal state, based on a common 3 d preceding-state window followed by 2 d persistence in the new state. In contrast, the stable boundary represents the final seasonally sustained transition satisfying the longer-term stability, terminal-anchor, and reversal criteria. When no distinct early event was detected, the early and stable boundaries coincided. The four possible ice-out/freeze-up combinations were therefore early/early, early/stable, stable/early, and stable/stable. The primary analysis used the stable/stable combination; the other combinations were evaluated only as boundary-definition sensitivity tests.

The corrected cross-year ice-season-duration trend based on stable ice-out and early freeze-up was −4.23 d decade^{-1} ($p = 0.048$). However, the other three boundary combinations did not reproduce this nominally significant result. We therefore retain it only as a sensitivity result and do not use it as evidence of an established long-term change. The former within-calendar-year duration metric is no longer used; all duration results are based on the continuous interval from freeze-up in year $y - 1$to ice-out in year y.

The proposed replacement for manuscript Figure 9, reproduced below, presents the complete-record trends and their sensitivity to the choice of early and stable boundaries.

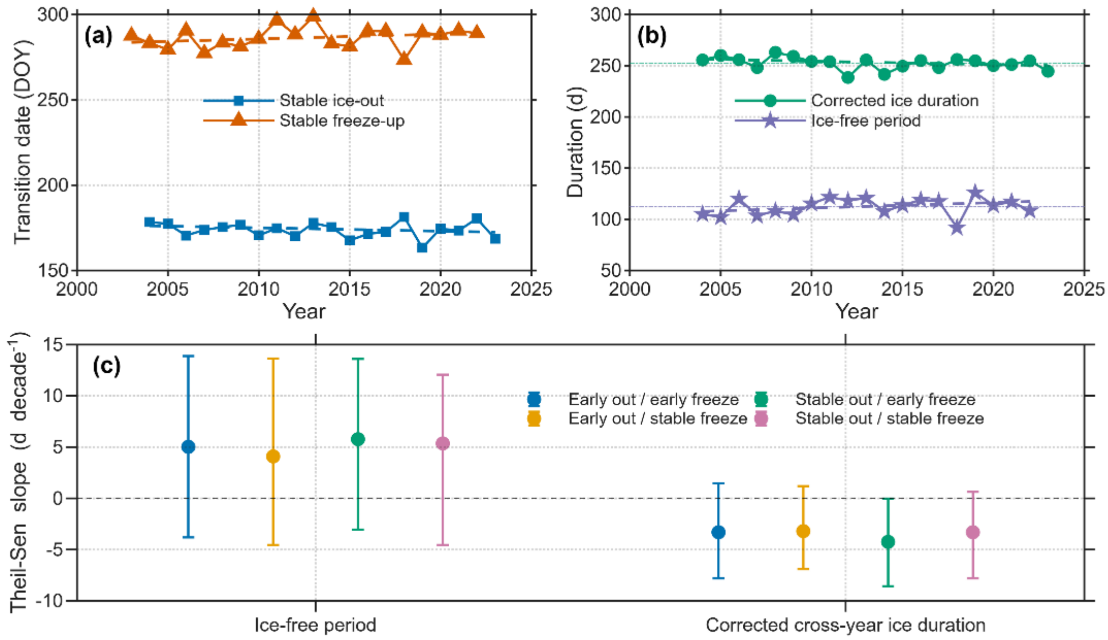


Figure 9. Long-term tendencies in GNSS-IR coastal sea-ice phenology at TUKT under the early- and stable-boundary framework. (a) Primary stable ice-out and stable freeze-up dates with Theil–Sen trends. (b) Ice-free period and corrected cross-year ice-season duration under the primary stable–stable definition. (c) Sensitivity of the duration slopes to four early/stable boundary combinations; bars show two-sided 95 % confidence intervals. Numerical results are reported in Table 4.

Changes to be incorporated in the revised manuscript: The title, Abstract, Sect. 4.2.2, Table 4, Figure 9, and Conclusions now retain 2018 and report complete-record estimates and uncertainty. The exclusion-based result and within-calendar-year duration have been removed from all analyses, figures, tables, and conclusions.

Major comment 8: Section 4.3: Interpretation of the thermodynamic mechanism

Status: Addressed

Reviewer comment. *The manuscript attributes the delayed freeze-up primarily to enhanced summer ocean heat uptake and a "thermal memory" effect. However, the study does not directly analyze ocean heat content or autumn ocean thermal anomalies sufficient to quantify this mechanism. Moreover, the freeze-up trend itself is not statistically significant either with or without 2018.*

The observed association with air temperature and FDD supports a thermodynamic influence on freeze-up timing, but it does not by itself demonstrate that enhanced summer ocean heat storage is the dominant mechanism responsible for the long-term trend. The authors should distinguish more clearly between mechanisms directly demonstrated by the present analysis and physically plausible interpretations based on previous literature.

Response. We agree. In a revised manuscript, we will no longer claim that enhanced summer ocean heat storage or a thermal-memory effect is demonstrated as the dominant cause of the freeze-up tendency. Air temperature and degree days are atmospheric thermal indicators, not observations of ocean heat content.

We also corrected for selection of the accumulation window using a max-statistic permutation test with 9,999 permutations, repeating the complete candidate-window search in every permutation. Stable ice-out was associated with TDD accumulated over DOY 84–173 ($r = -0.817$; $R^2 = 0.668$; search-adjusted $p = 0.0004$), and stable freeze-up with FDD over DOY 260–289 ($r = -0.877$; $R^2 = 0.769$; adjusted $p = 0.0001$). These results support thermodynamic associations with interannual transition timing but do not identify a unique causal mechanism for the long-term tendencies.

The discussion now also recognizes wind-driven ice motion, advection, storms, snow conditions, fracturing, and other mechanical processes. The two optimized correlations were not formally compared, and the non-significant freeze-up trend is stated explicitly.

Changes to be incorporated in the revised manuscript: The search procedure and max-statistic adjustment will be added to Sect. 3.3.4. Sections 4.3.1–4.3.2, Figure 11, and the Conclusions now distinguish the demonstrated degree-day associations from plausible mechanisms discussed in the literature and remove causal ocean-heat-storage language.

Major comment 9: Section 4.4: Interpretation of the 5.5-day GNSS-IR–IMS offset

Status: Addressed

Reviewer comment. *The reported mean freeze-up offset of 5.5 ± 3.7 days is potentially interesting, but its physical interpretation depends strongly on the IMS spatial aggregation described in Section 2.3.2. If the IMS phenological date is actually derived from an approximately 44×44 km window, it is not appropriate to explain the offset simply as a comparison between the GNSS-IR footprint and a single 4-km or 16-km2 IMS pixel.*

Therefore, until the exact IMS processing and spatial support are clarified, the 5.5-day lead should not be interpreted as definitive evidence of a sensor-resolution effect. The distinction between spatial-resolution effects, different spatial sampling domains, and genuine nearshore–offshore differences should be made more carefully.

Response. We agree. The IMS spatial support is now the four-cell neighbourhood described in Major Comment 2, and all comparison statistics have been recalculated consistently. For 2004–2022, stable freeze-up showed strong correspondence ($r = 0.845$; $p < 0.001$; $n = 19$). The mean GNSS-IR-minus-IMS difference was −3.22 ± 3.41 d, the

RMSE was 4.63 d, and GNSS-IR was earlier in 15/19 years.

Ice-out correspondence was weak (mean difference −0.58 ± 7.45 d; RMSE 7.27 d; r = 0.048; p = 0.846; n = 19), and corrected-duration correspondence for 2005–2022 was also weak (mean difference +3.15 ± 8.27 d; RMSE 8.64 d; r = 0.207; p = 0.411; n = 18).

In a revised manuscript, we will not assign the freeze-up difference to spatial resolution alone. Possible contributors include the distinct local and four-cell footprints, sensitivity to partial or newly formed ice, and spatial variability in freeze-up timing, but their contributions cannot be partitioned. The result is not presented as evidence of frazil-ice detection or a uniquely identified nearshore–offshore mechanism.

The proposed revised cross-scale comparison is reproduced below as Figure 12.

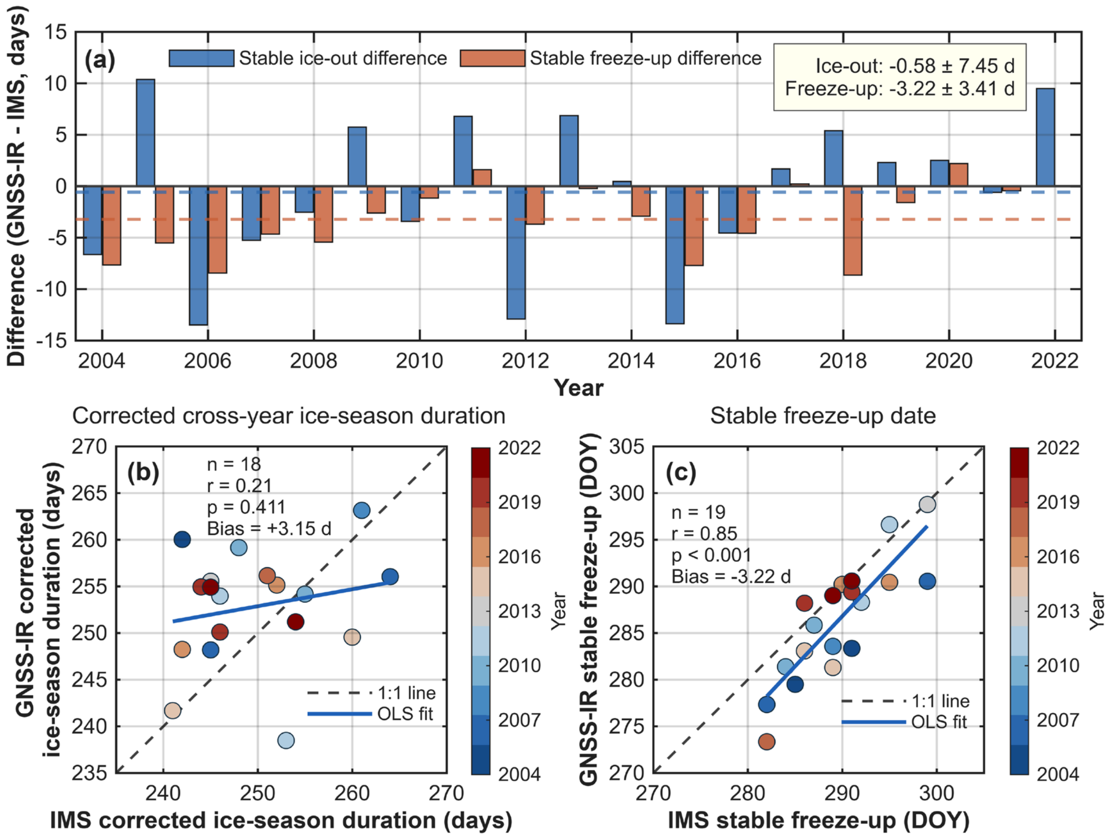


Figure 12. Cross-scale comparison of GNSS-IR and IMS sea-ice phenology. (a) Annual transition-date differences (GNSS-IR minus IMS) for 2004–2022; dashed lines are means. (b) Corrected cross-year ice-season duration for 2005–2022. (c) Stable freeze-up for 2004–2022. In (b–c), colours denote year, dashed black lines show the 1:1 relation, and blue lines show ordinary least-squares fits.

Changes to be incorporated in the revised manuscript: The IMS definition has been corrected in Sect. 2.3.2. Section 4.4, Table 5, Figure 12, and the Conclusions will report the recalculated cross-scale results and replace the former single-mechanism explanation with an explicitly non-causal interpretation.

Minor Comments

Minor comment 1: Lines 55 and 57

> **Reviewer comment.** *Lines 55 and 57:The manuscript contains several formatting and language issues, including duplicated punctuation and inappropriate spacing. A careful proofreading of the manuscript is recommended.*

Response. Thank you. The identified errors were corrected, and a manuscript-wide editing pass removed duplicated punctuation, inconsistent spacing, repeated text, and additional grammar problems. Terminology, hyphenation, units, p-value wording, and figure and table references were standardized.

Changes to be incorporated in the revised manuscript: Corrections have been made throughout the Abstract, main text, captions, tables, and references.

Minor comment 2: Line 63

> **Reviewer comment.** *Line 63:Unless the dielectric permittivity values are representative of conditions at the study site or their applicability to the TUKT environment can be justified, I suggest avoiding overly specific numerical values here.*

Response. We agree. The values are no longer presented as measurements representative of TUKT. They are retained only as explicit assumptions for the idealized co-polar RHCP calculation in Figure 3, with the limitation that sea-ice and seawater permittivities vary with temperature, salinity, brine volume, and surface state. The observed-profile analysis in Figure 4 now provides the empirical test of the expected contrast.

Changes to be incorporated in the revised manuscript: The assumptions and limitations will be added to Sect. 3.1.2 and the Figure 3 discussion; the direct observed comparison will be added to Sect. 4.1 and Figure 4.

Minor comment 3: Figure 1

> **Reviewer comment.** *Figure 1:The station symbols in the legends do not appear to be consistent. In addition, Figure 1b appears to show that some of the plotted First Fresnel Zones, particularly those associated with the 5° and 10° elevation angles, overlap the land surface. Please verify the reflection-zone geometry and the statement that the selected footprints are entirely located over water.*

Response. We agree. The station marker and label are now consistent, and redundant legend boxes have been removed. The Fresnel ellipses were recalculated using the verified L1 geometry described in Major Comment 1. The manuscript no longer states that every footprint lies entirely over water and explicitly acknowledges possible limited contributions from the immediate coastal margin.

Changes to be incorporated in the revised manuscript: Figure 1, its caption, and Sect. 2.2 will be revised consistently.

Minor comment 4: Table 1

> **Reviewer comment.** *Table 1:Describing Sentinel-2 as having a temporal resolution of 5 days may be misleading for this application. Although the nominal revisit interval may be approximately 5 days, the number of usable cloud-free images over the study area is much lower. Please distinguish nominal revisit time from actual image availability.*

Response. We agree. In a revised manuscript, Table 1 will no longer describe Sentinel-2 as providing five-day temporal resolution for this application. Instead, it will identify five selected cloud-free optical scenes from 2019: three Sentinel-2A scenes and two Landsat 8 scenes. Section 2.3.4 will list their exact acquisition dates and clarify that the imagery provides qualitative event-scale context rather than continuous temporal coverage.

Changes to be incorporated in the revised manuscript: The optical-satellite imagery entry in Table 1 and Sect. 2.3.4 will be revised accordingly.

Minor comment 5: Figure 3

Reviewer comment. *Figure 3:Figure 3 should be placed after its first citation in the main text.*

Response. Thank you. Figure 3 has been moved to follow its first citation. The text now introduces the modeled decomposition immediately before the figure, with its numerical and physical interpretation following in Sect. 3.1.2.

Changes to be incorporated in the revised manuscript: Figure 3 has been repositioned after its first citation, and the surrounding text has been reordered accordingly.

Minor comment 6: Figures 4 and 7

Reviewer comment. *Figures 4 and 7:Please define the meaning of the ±1σ uncertainty envelope. It is unclear whether it represents inter-satellite variability, uncertainty in the filtered estimate, a confidence interval, or another quantity.*

Response. We agree that the original captions were unclear. For the short-term and complete-record AIF figures, residuals are individual reference-centred AIF observations minus the filtered AIF. At each observation time, the local standard deviation is calculated from all finite residuals in a centred 24 h window. The envelope is the filtered AIF ±1 local residual standard deviation and represents short-timescale scatter; it is not a confidence interval, transition-date uncertainty, or purely inter-satellite variability. The shaded bands in Figure 4 are separately defined as pointwise 95 % bootstrap intervals across annual profiles.

Changes to be incorporated in the revised manuscript: The residual-envelope calculation has been defined in Sects. 3.2 and 3.3.1 and in the captions of Figures 5 and 8; the distinct bootstrap intervals in Figure 4 are defined separately.

Minor comment 7: Figure 6

Reviewer comment. *Figure 6:The Sentinel-2 images shown in Figure 6 appear not to match the imagery available for the corresponding dates and location in Copernicus Data Space. The authors are requested to verify the acquisition dates and image products.*

Response. We thank the reviewer for identifying this issue. We rechecked the original image files and confirmed that the acquisition dates were correct, but the original caption incorrectly described all of the optical images as Sentinel-2 products. The figure was constructed from five independently acquired scenes: three Sentinel-2A MSI Level-2A products and two Landsat 8 Level-1 precision- and terrain-corrected products. Their verified product identifiers and acquisition information are as follows:

Date	DOY	Sensor and product	Product identifier
8 June 2019	159	Sentinel-2A MSI Level-2A	S2A_MSIL2A_20190608T210021_N0500_R100_T08WNC_20230628T150333.SAFE
9 June 2019	160	Landsat 8	LC08_L1TP_063011_20190609_20200828_02_T1
15 October 2019	288	Landsat 8	LC08_L1TP_063011_20191015_20200825_02_T1
16 October 2019	289	Sentinel-2A MSI Level-2A	S2A_MSIL2A_20191016T210451_N0500_R100_T08WNC_20230615T130429.SAFE
20 October 2019	293	Sentinel-2A MSI Level-2A	S2A_MSIL2A_20191020T204501_N0500_R014_T08WNC_20230526T205350.SAFE

The corresponding Sentinel-2 sensing times were 21:00:21 UTC on 8 June, 21:04:51 UTC on 16 October, and 20:45:01 UTC on 20 October 2019. The later 2023 timestamps in the Sentinel-2 product identifiers denote product-generation or reprocessing times, not image-acquisition dates. Similarly, the 2020 fields in the Landsat identifiers are processing dates; the acquisition dates are the preceding 2019 fields. The two Landsat products both correspond to path 063 and row 011.

The corrected 3 × 2 figure contains six panels derived from these five scenes. Panel (a) uses the Sentinel-2A image from 8 June; panels (b) and (c) use the same Landsat 8 image from 9 June, displayed as a local view and a regional overview, respectively; panel (d) uses the Landsat 8 image from 15 October; and panels (e) and (f) use the Sentinel-2A images from 16 and 20 October, respectively. The regional overview includes a rectangle identifying the local area, and the TUKT station is marked consistently in all panels.

These optical images are used only to provide qualitative spatial context for the surface conditions around TUKT during the spring and autumn transition periods. They are not used to calculate the AIF transition dates, estimate trend magnitudes, or provide independent quantitative validation.

Changes to be incorporated in the revised manuscript: The original Figure 6 will be replaced by the corrected Figure 7. Section 2.3.4 and the Figure 7 caption will identify the sensor, acquisition date, and product associated with each panel. Section 4.1 will clarify that the Sentinel-2 and Landsat 8 images provide qualitative spatial context only and are not used in the quantitative phenology or trend calculations.

Minor comment 8: Tables 3 and 4, Figure 8, and associated text

Reviewer comment. *Tables 3 and 4, Figure 8, and associated text:The manuscript uses several different temporal ranges, including 2004–2022, 2004–2023, a 19-year record, and a 20-year record. Please make these descriptions consistent and state the exact number of observations used in each regression.*

Response. We agree and now distinguish the overall archive span from the sample used for each metric. The record extends from late August 2003 to early September 2023. Stable ice-out uses 2004–2023 (n = 20), stable freeze-up 2003–2022 (n = 20), ice-free period and matched-year contrast 2004–2022 (n = 19), and corrected cross-year ice-season duration 2004–2023 (n = 20). GNSS-IR–IMS transition comparisons use 2004–2022 (n = 19), and the corrected-duration comparison uses 2005–2022 (n = 18). Missing endpoint events are not treated as detector failures.

Changes to be incorporated in the revised manuscript: Periods and sample sizes have been reconciled in Sects. 2.2, 3.3.3, and 4.2–4.5, Tables 3–6, Figures 8–12, and the Conclusions. Each regression table and caption now identifies its exact period and n.

We thank the reviewer for the constructive comments. Our final responses distinguish direct evidence from physically plausible interpretation and document the additional analyses completed in response to the review. These include clarification of the GNSS-IR and IMS spatial supports, evaluation of the observed ice–water AIF contrast and AIF–AMP performance, quantification of receiver-period effects, reanalysis using symmetric gap-aware transition criteria, retention of 2018 in all primary analyses, and explicit reporting of uncertainty estimates. 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 transparency, methodological consistency, and reproducibility of the study.