General Comments:

This is an extremely thorough and comprehensive analysis of whether Early Warning Signals (EWS) are present before Dansgaard-Oeschger (DO) events in Greenland ice cores. Importantly, this study rigorously addresses the methodological limitations and discrepancies that have led to conflicting results in previous studies on this topic. The statistical approach used to detect EWS is justified by the close match between the analytical and numerical distributions for the number of false positives. Although this detection of significant EWS preceding individual transitions is carried out extremely well, I believe there is an issue with the analysis of whether the number of observed EWS is in turn significant itself. This could impact the findings and therefore needs to be carefully addressed. Overall, I recommend that this manuscript is published subject to revisions.

We thank John Slattery for the very helpful and thorough review. The comments will be addressed point-by-point below.

Specific Comments:

The only major issue in this otherwise excellent manuscript concerns the analysis of whether the number of observed EWS is significant. As the authors correctly state on line 218: "*For $x \sim B(17, 0.05)$, it is $P(x \leq 2) \approx 0.9497 < 0.95$ and $P(x \leq 3) \approx 0.9912 > 0.95$.*" However, the authors then mistakenly infer from this that "*at a confidence level of 95%, we expect at most two events to show spurious significant early warning, and observing three significant EWS is statistically significant*." In fact, the number of EWS required for statistical significance at the 95% level is N, where N is the smallest integer such that, for $x \sim B(17, 0.05)$, $P(x < N) > 0.95$. The crucial difference is that the probability of x being less than **but not equal to** N must be more than 95%, not less than or equal to as the authors imply. The significance threshold at the 95% confidence level using this analytical distribution is therefore four significant EWS observed, not three.

One can consider this in an equivalent way that may be clearer by thinking instead about the p-value as compared to the significance level (i.e. 1 - confidence level). A result is significant at the 5% significance level if, under the null hypothesis, the probability p of observing a result at least this extreme is less than 5% (i.e. $p < 0.05$). In our case, the number of observed EWS required for significance is N, where N is the smallest integer such that, for $x \sim B(17, 0.05)$, $P(x \geq N) < 0.05$. $P(x \geq 3) = 1 - 0.9497 = 0.0503 > 0.05$, and so observing three EWS is not quite statistically significant at the 5% / 95% level, whilst observing four is.

To see clearly that the authors' approach is mistaken, consider Figure 3(a&b). Both the analytical and numerical distributions show that there is a 16% probability of 2 out of the 17 transitions showing false positive EWS. Despite this, the authors indicate in 3b that observing two EWS is significant at the 95% level using the numerical distribution. Elsewhere, including Figure 10, they also indicate that observing two EWS is significant at the 90% level for both distributions. Observing two EWS cannot be significant at either

confidence level or with either distribution, though, because this happens by chance 16% of the time! For another example, consider the distribution for simultaneous EWS in Figure 3c. Using the authors' logic, $P(X \geq 0) > 0.95$ and so 0 transitions with simultaneous EWS in both indicators would be a significant positive result, which clearly cannot be the case. I hope that these examples demonstrates that my comment here is not merely a statistical foible or a petty criticism, but that it has a real impact on the findings of this study.

We thank the referee for pointing this out and agree that the approach we used to calculate significance thresholds for the number of significant EWS was indeed wrong. The statement in lines 218-220 will be corrected accordingly and the significance thresholds will be corrected in the revised manuscript and all relevant figures (i.e. Figures 3, 10, A3, S14, S16, and S19).

The comparison of the analytical and numerical distributions is a fantastic way to show that the significance test for EWS preceding individual transitions works as intended, and I applaud the authors for including this. However, having done so, I think it would be better to then consider only the significance threshold for the number of observed EWS derived from the analytical distribution. This would simplify the analysis by making the threshold the same for all records and indicators. Currently there is sometimes (e.g. Figure 3b) a discrepancy between the thresholds for the two distributions, even though they match very well, just because $P(x < 3)$ is so incredibly close to 0.95.

We agree that the analysis would be simplified by only using the analytical significance threshold. Hence, we will remove references to the numerical distribution when comparing the different records and summarising the results.
Nevertheless, we note that numerical distributions have only been calculated for the NGRIP record with 5-year resolution, and argue that the different significant thresholds stemming from the numerical and analytical distributions are worth mentioning in this case. Thus, they will only be mentioned briefly in the main text of Sections 2.4 (Methods - Expected number of spurious significant EWS) and 3.7 (Results) in the revised version of the manuscript.

Line by line comments:

Line 22 and elsewhere: This study describes $\delta^{18}O$ in Greenland ice cores as a local temperature proxy, following the traditional interpretation. However, recent isotope-enabled modelling (Buizert et al. 2024, https://doi.org/10.1073/pnas.2402637121) suggests that winter sea ice variation may instead be the dominant control on $\delta^{18}O$ during DO events. I suggest that this new interpretation should be briefly discussed, either in the introduction or in Section 4.2.

We greatly appreciate this input and will include a short discussion on this in Section 4.

Figure 3: I think it would also be better to place the significance threshold lines between integers, as it is currently unclear whether observing a number of EWS equal to the significance threshold is significant or not. Indeed, Figure 3c seems to suggest that the

significance threshold is 0, if interpreted in the same way as a & b, which cannot be the case. This should of course also account for the corrected significance thresholds based on my main comment, and the same also applies to Figures A3 and S19.

We agree and will place the (corrected) significance threshold lines between integers in Figures 3, A3, S16 and S19.

Lines 253-254: "*Furthermore, we don't restrict the search for wavelet-based EWS to the GS until 200 years prior to events to include potential influences of the transitions themselves.*" This sentence is unclear to me.

This sentence will be changed to "*Furthermore, we don't restrict the search for wavelet-based EWS to the GS until 200 years prior to events, as in Boers (2018), to exclude potential influences of the transitions themselves. Instead …*"

Lines 287-288: "*Though, observing two significant EWS in α_1 is only significant with respect to the analytical, but not the numerical null-distribution.*" Based on Figure 3b this appears to be the wrong way round, as the numerical threshold is two EWS whilst the analytical threshold is three. Either way, as mentioned above, I think it would simplify the analysis to consider only the analytical null distribution.

We agree that this is the wrong way round. Following the comment above, this statement will be corrected, including only the corrected analytical threshold.

Figure 10: It is difficult to distinguish between zero and undefined using this colour scheme. The circles indicating significance should also be corrected as discussed above.

We will change the colour scheme and adjust the significance thresholds.

Lines 480-481: "*Recent advancements in EWS methods have … introduced new methodologies (Clark et al., 2002)*." It seems odd to call a study from 23 years ago a recent advancement. Perhaps a different reference was intended here, otherwise this sentence should be reworded slightly.

We agree. A different reference was intended here indeed. It will be replaced with (Clarke et al., 2023, https://doi.org/10.1088/1748-9326/acbc8d).

Lines 529-530: "*It has been shown before that the current NGRIP site was located at a higher altitude and further upstream, closer to NGRIP than it is today.*" This sentence is unclear. I think that the authors perhaps intended to write NEEM here instead of NGRIP.

We indeed intended to write NEEM instead of NGRIP and will correct this sentence to "*It has been shown before that the current NEEM site was located at a higher altitude and further upstream, closer to NGRIP than it is today [...].*"

Fig S3c in Supplementary Information: The line for the 95% confidence interval is either hidden or missing.

We will recreate this figure, including the hidden line for the 95% confidence interval.

Technical Comments:

We appreciate all the technical comments below and will correct these mistakes before resubmission.

Line 202: *"allows to handle data"* is missing a word. This should perhaps read "allows us to handle data".

Line 287: *"…and autocorrelation for DO-12 Though, observing two significant EWS…".* I think there ought to be a full stop between *"DO-12"* and *"Though"*.

Line 314: *"resolutions.Another"* is missing a space after the full stop

Figure 7 caption: *"(e-f) Same as (c-d) but with modified estimator calculation.(g-h) Same as (e-f) but with modified data preprocessing.Line colours and shadings are applied in the same way as in Fig. 4."* Spaces are missing after both full stops.

Line 459: *"notably DO-1,6"* is missing a space after the comma.

Lines 519-520: *"(Guillevic et al., 2013; Seierstad et al., 2014; Capron et al., 2021; Steen-Larsen et al., 2013)"* These references are not in chronological order.

Line 548: *"on parts the record"* is missing a word.

I hope that these comments are helpful, and I look forward to reading the authors' response.

We are grateful for these detailed and valuable comments and thank John Slattery for his input and feedback.