This paper presents a thorough analysis of Early Warning Signals (EWS) prior to the abrupt Dansgaard-Oeschger events observed in Greenland ice-core records. All the available deep records, GRIP, GISP2, NGRIP and NEEM are used for the analysis. EWS are changes in statistical properties of a time series indicating a bifurcation-induced transition (b-tipping), they will not appear prior to a noise-induced transition (n-tipping). The aim is thus to identify for each of 17 DO-events in the well-dated past 60kyr records which would be due to b-tipping and which would be due to n-tipping assuming a classical bistable dynamics. As the detailed dynamics of the transitions are largely unknown, the simplest assumption (Occam's razor type of argument) is that of a saddle-node bifurcation in a system subject to noise. In such a system variance will, from the fluctuation-dissipation theorem, increase when approaching the bifurcation point, likewise will the autocorrelation increase. This is the phenomenon of critical slow down. For any other suggested scenarios for the transitions, different EWS could potentially be detected. Since the transitions documented in the paleoclimatic records have already happened, detected EWSs obviously play the roles of hindcasts rather than forecasts, thus the purpose of detecting EWSs is rather dynamical system identification.

A fair statistical significance test is constructed by booth-strapping through generation of so-called Truncated Fourier Transform Surrogates (TFTS), which is just surrogate timeseries constructed by randomly choosing phases (not shuffling) of the Fourier-coefficients while keeping the amplitudes of the original signal. "Truncated" refers to not changing phases of the long wavelength coefficients to preserve trends in the timeseries. Since the variance and the autocorrelation in a time series only depends on the amplitudes of the Fourier coefficients, the TFTS will have the same variance and autocorrelation as the original time series over the full glacial state (GS) period analyzed. The EWS indicators are now calculated within 200y running windows for each of the GS periods prior to the DO-transitions and the slope of the linear fit of this indicator time series is calculated and a significant slope (at the 95% confidence level) is identified from the distribution of slopes in the TFTS time series. From this analysis it is established that only a few DO-events are preceded by EWS, in agreement with the expectation that about one of the 17 DO events should be significant at the 95% confidence level, motivating the title of the paper.

The findings confirm our earlier findings (Ditlevsen and Johnsen, 2010), so in some sense this is a reporting of negative results. However, I find that the paper presents useful methods for this kind of analysis, thus I recommend publication. I do though recommend a revision for clarifications and better readability:

We thank Peter Ditlevsen for this helpful and thoughtful review. All comments are addressed in our point-by-point responses below.

1. The GS vary in duration, a typical GS lasts perhaps 2ky, which means that there are ten independent 200yr window measurements. Thus, the linear trend is made for only 10 points or maybe even less. Furthermore, for the 20yr resolution records, there are only ten points within a 200yr window, from which the EWS are

calculated. A discussion of the uncertainty and the quality of the estimates is lacking.

We agree that the individual stadials provide rather short time series to conduct EWS analyses on. While there are indeed only few independent, non-overlapping window measurements, we note that we use sliding windows to compute the CSD EWS indicators. So, for a typical stadial of 2ky, the linear trend of EWS indicators is calculated from 360 (180, 90) data points for records in 5- (10-, 20-) year resolution.
By comparing records from different ice core locations with varying temporal resolutions, as well as various methodological choices, our analysis shows that the EWS indicators used here are indeed sensitive to these factors. As such, this gives an insight into the uncertainties of EWS indicators and provides a more comprehensive view than previous studies on EWS preceding DO events. We think that the uncertainty and the quality of the EWS estimates is carried out in detail regarding the results of the significance tests, both with respect to individual trends, and with respect to the number of significant trends. Also note that the significance test for the trends is based on phase surrogates of the underlying stadial sections and thus incorporates the length of these time series segments. We will address this in the revised manuscript.

2. A consistency check between significant EWSs found for some, but not both EWS and some, but not all records (which are obviously false positives) and the number of false positives expected from the boot-strapping should be made.

   We present the numbers of significant EWS for individual indicators and both occurring simultaneously for all ice core records. Further, we also constructed null-distributions for the number of false positives for individually and simultaneously increasing indicators, which indicate how many indicator increases are expected to occur by chance at different significance levels in a given record.
   Nevertheless, because the records considered differ in many aspects, such as ice core location, processing and temporal resolution, we believe that observing EWS in some, but not all records does not necessarily imply a false positive. It could simply be that an underlying true EWS is masked in an individual core, with preprocessing steps affecting the different EWS indicators in different ways
   In order to ease the comparison, we will include a figure or table showing which DO event is preceded by which combination of EWS for each record and will comment more on this in the revised manuscript.

3. Figures 4-9 are difficult to read, unless they are only to be read like white-pink-red barcodes. Consider showing just one full width time series (for each EWS) and present the consistency between resolutions/methods/records in a figure similar to Figure S7 or S8, in order to get a better overview. There is a lot of information in the text, which makes reading difficult.

We agree that Figures 4-9 and the main text contain a lot of information. For the revised manuscript, we will also show our results in a more aggregated way to facilitate readability. Some of the current plots showing indicator time series will be moved to the supplementary material.

4. Increase in the weights of specific wavelet coefficients and increases in local Hurst exponent, H, are suggested as EWS. However, it is not argued what the assumed underlying (complex) system exhibiting these EWSs are. There are references to earlier papers by the same authors (Rypdal, 2016, Boers, 2018), but these references do not provide such justifications. It is mentioned that the Hurst exponent is an estimate for the correlation, but if this is the only argument for calculating H, the authors should at least argue why it is more reliable to calculate H, than just analyze the autocorrelation (which is also done). It would substantially strengthen the paper if such arguments could be presented.

   We agree that the presentation in our paper should be more self-contained, and will summarize the main reasons why the Hurst exponent is indeed a useful scale-aware extension of the AC1 coefficient in the revised version. We will explain why H is a scale-aware measure of the memory in the time series, which increases as a bifurcation is approached, motivating its use as an EWS. Moreover, since H has been used in previous papers, we believe it is important to include it here as well.

5. A discussion of how a Hurst exponent can meaningfully be calculated from about one decade of time scales is lacking. What are the uncertainties?

   We not 100% sure we understand the referee’s concern here, but will clarify how exactly we compute the *local* Hurst exponent in our study; moreover, uncertainties in the estimation of H are captured by our two significance tests, first based on phase surrogates to test individual trends, and second for the number of significant trends in the full time series.

6. In line 230 it is stated that for a linear stochastic process increase in variance and increase in autocorrelation are independent. This is not true: For the OU process x, we have Var(x) ~ -1/(log AC(1)).

   We agree, of course. Nevertheless, the *estimates* of increases in variance and autocorrelation are independent under the null hypothesis that there are no parameter changes in the underlying system. This sentence will be reformulated for clarity accordingly.

7. The notation sigma^2 for Var(x) is unfortunate, since the underlying assumptions of the EWS is that the locally stationary process is the OU process: dx = -alpha x dt + sigma dB, where alpha_1 = alpha dt (lag-1) AND Var(x)=sigma^2/(2 alpha).

Thus sigma^2 represents the square of the intensity of the noise. I recommend using Var(x) for the variance.

We agree and the notation will be changed to Var(x) in the revised manuscript.

I hope these comments are useful for the authors.

We highly appreciate these comments and thank Peter Ditlevsen for his input and feedback.