Reply to Referee 2

First of all, thank you very much for reviewing our manuscript in detail and giving us very valuable feedback. In what follows, we respond to your comments and questions, point by point, and propose changes to the manuscript in accordance. We think that these changes will improve the quality and clarity of our manuscript.

In order to improve the readability of our replies we applied a color/type coding to discriminate our replies from the referee's comments. We have attached our replies as a pdf document since color coding is not available in the browser-based text editor.

Color/type coding:
Comment by the referee.
Reply from the authors.

I was sad to see the code used in this paper was not shared 'by default'.

We are sorry for that. We will make our codes available by default in the revised manuscript. Generally, we prefer to publish code only upon acceptance of the manuscript but of course we are happy to share the code with the referees already beforehand.

There are lots of time series here: 2 proxy variables, 2 early warning signals, 3 cores, 12 interstadials, 5 window sizes and 6 smoothing spans leading to 3480 analysed time series (when accounting for the fact that not all cores have all interstadials in them), although relatively few give SPS (31%). If there is a common mechanism at work, why is this the case?

As mentioned in the manuscript, robust statistical precursor signals (SPS) are more likely to be observed in longer interstadials than in shorter ones (cf. Figs 4c and 4d). We speculate that it happens because, for short interstadials, it is difficult to detect statistical changes in the stability. In the revised manuscript, we provide such a numerical example as a supplementary material and will expand the discussion section to make this clearer.

Can the authors be sure the results are not down to chance? The authors argue that more SPS are observed than would be expected by chance, but none of the time series for each interstadial are truly independent, and there being a false positive in one time series would increase the chance of there being a false positive in another.

As you point out, the time series of different proxies and cores are not independent if they are observations from the same DO event. Thus, whether the significant SPS are obtained by chance or not should be assessed proxy by proxy and core by core, that is, row by row in Fig. 4d. Even if we assess the results in this way, the risk of obtaining the results by chance is quite low. For example, for the case of the variance of NGRIP d18O, we observed 6 robust SPS over 12 interstadials, which is extremely unlikely to be obtained by chance. We have obtained no robust SPS for the case of the autocorrelation of GISP2 d18O. This is the only case where we cannot reject the null hypothesis.

Furthermore, looking at figure 4d, different cores give different results for the same interstadial, e.g. in GI-20 only 4 cores give any SPS and only one 'robustly'. Different interstadials give different amounts of SPS, for example GI-14 gives many robust SPS but GI-19.2 doesn't. How do the authors account for this?

Thank you for raising this question. Whether the robust SPS is observed or not depends on the proxy and core. This is possibly because different proxies from different cores are contaminated by different types and magnitudes of noise. For example, d18O may record local temperature fluctuations and Ca2+ may record turbulent fluctuations of local wind circulations. We will comment on this point in the revised manuscript.

When looking for SPS, the time series must be decomposed into an slowly changing equilibrium state and fluctuations about that state. As a lot of the signal in this case for SPS comes from 'rebound events' the authors are assuming that the rebound events represent fluctuations rather than changes in the equilibrium state. What is the justification for this?

We assume that the rebound events present fluctuations due to loss of stability of quasi-equilibrium states, given that the four low-dimensional dynamical systems in Fig. 5 can qualitatively mimic the rebound events. The alternative assumption of rebound events as intermediate equilibrium states is mentioned in the discussion section, citing Lohmann et al. The justification of our assumption is beyond the scope of this work. We will more explicitly note this assumption in the revised manuscript.

The authors may want to discuss mechanisms that can lead to changes in variance and autocorrelation not due to changing stability but due other factors. For example, due to changes in the properties of the climate forcing.

Thank you for this comment. In the revised manuscript, we will mention that the increases in variance toward a DO cooling transition may be explained by

the larger climatic fluctuations in colder states, but the increases in autocorrelation cannot generally be explained by it. We will nevertheless add a discussion on the possibility of false positive and negative SPS.

Furthermore, changes in the statistical properties in the measurement process may also affect the results. For example, measurements in the ice cores further in the past may be more uncertain and therefore noisier, but measurements closer to the present may be less noisy and therefore more correlated.

We fully agree that the uncertainty of the data is higher in the older part of the cores. On the other hand, we don't see systematic changes in the results detecting SPS (e.g., Fig. 4d).

Specific Comments:

Line 11: Should be rebound events not rebound event

Corrected.

Line 15: This tipping definition excludes N-tipping, which has no thresholds. Different authors define tipping differently but as there is disagreement over whether DO events are N or B tipping I wonder if it is better to adopt a definition compatible with the Ashwin 2012 typology?

Thank you for this comment. Yes, the description of "tipping points" in line 15 was restricted to that of bifurcation-induced tipping points. We will adopt a wider description in the first paragraph such as "a threshold crossed irreversibly by the system's dynamics". Then in the second paragraph, we will mention B-, N-, and R-tipping citing Ashwin et al. and others.

Line 91 "R^2 = 0.95", what fit is this measuring?

R^2 = 0.95 is the coefficient of determination for the correlation between the length of rebound event and the length of interstadial. In the revised manuscript, we have rewritten this and now say that their durations are correlated with R^2=0.95.

Line 114: The autocorrelation is different to that in Bury et al who have C(tau) = (cos (omega tau)) exp(mu \|tau\|)

Thank you. Corrected!

Line 115: Should be "increase or decrease"

Corrected.

Line 117: How do the authors know tau is sufficiently small, especially as omega may also be changing?

In theory, we can calculate the autocorrelation function over the running window and thus can choose a sufficiently small tau. Here, the minimum sampling time is taken. The frequency omega itself does not change across the Hopf bifurcation (Strogatz 2018).

Line 120: Is a linear fit suitable if half of the interstadial is used i.e. 500+ years? Could the stable state be changing nonlinearly in this period?

The locally estimated scatterplot smoothing (LOESS) used in this study performs a local polynomial (here simply linear) fit in its procedure, giving more weight to points near the point whose response is being estimated and less weight to points further away. Thus, it can provide smoothed series for time series with nonlinear trend even if the local fit is linear. We will explain this point in detail in the revised manuscript.

Line 187: makes reference to interstadials shorter than 1000 years but Line 106 implies the authors are excluding interstadials shorter than 1000 years. Have I misunderstood?

Sorry the sentence was misleading. In section 3.1, we have analyzed interstadials longer than 1000 years, but we have examined high-resolution interstadial data longer than 300 years in section 3.2. Thus the data between 300 and 1000 years is actually included in section 3.2. In order to avoid confusion, we simply say that “Robust SPS have not been observed for short interstadials again.”

Line 295-298: "can be shown to be 0.05". I think it would be helpful to show this. When I run the authors code I do not get any output like 0.044 or 0.042, but I may be running the code incorrectly. Is this calculation included in the shared code?

Excuse me, we have included the codes for main figures, but some of the codes for appendices and supplementary files are not included because some of them are tedious. However, we will upload all of the codes to a repository when submitting the revision.

Figure 4d: Could the colormap used in this figure be changed to a diverging colormap, with its centre at 15, so that it is easy to see if an SPS is robust. Currently

it is difficult to know if the colours correspond to values larger than or smaller than 15.

Thank you for this comment. We will improve the Figure 4d so that we can know if the corresponding values are above or below 15.