**Reply to Reviewer Dr Michel Crucifix**

We are grateful to the Reviewer for the comments that helped us improve the manuscript.

- *The problem with this method is that it relies on two hypotheses that may be unmet in the present application: stationarity over the analysis window (the $\partial/\partial t$ term is dropped in equation (3)), and additive and stationary Gaussian noise (otherwise additional terms would appear in equation (3), invalidating the solution (4)). Stationarity is particularly problematic during the deglaciation. The authors claim to identify four distinct states, which they admittedly cautiously interpret as representative of the full-glacial, Younger Dryas, Bölling-Alleröd and Holocene states. However, the residence time in these different states is large enough compared to the deglaciation time scale to call into question the stationarity hypothesis.*

  We agree that the stationarity of the ice-core records is problematic, especially through the deglaciation. We have reduced/removed discussion of the 4 states during deglaciation, noting instead that the data may represent 2 states around a moving trend.

  During the time interval 60–20 kyr BP the case for stationarity is stronger; we are therefore confident that our main result (bifurcation from two states to one state by 25 kyr BP) is robust.

- *Second, Ditlevsen (1999) identified a strong alpha-stable noise component in the Greenland Ca record, that may interfere with the robustness of the state identification algorithm because it alters the form of the Fokker-Planck equation. A suggestion would be to test the method with assumptions (3) and (4), but using surrogate data including an alpha-stable noise component.*

  We have tested the original method by applying it to the GRIP Ca data, and we have added a new figure 4 with the results.

  Comparing NGRIP $\delta^{18}O$ and GRIP calcium that was studied in Ditlevsen (1999), we first note that that Ca series has annual temporal resolution, whereas NGRIP temperature proxy that we study has 20yr resolution. To compare them on the same time scale, we show in the following figure both series, plotting NGRIP$\delta$, NGRIP$\delta$ ending at 10kyr BP (because GRIP Ca ends at 11kyr), and GRIP Ca starting at 60kyr BP (although the series is available from 91 kyr BP).

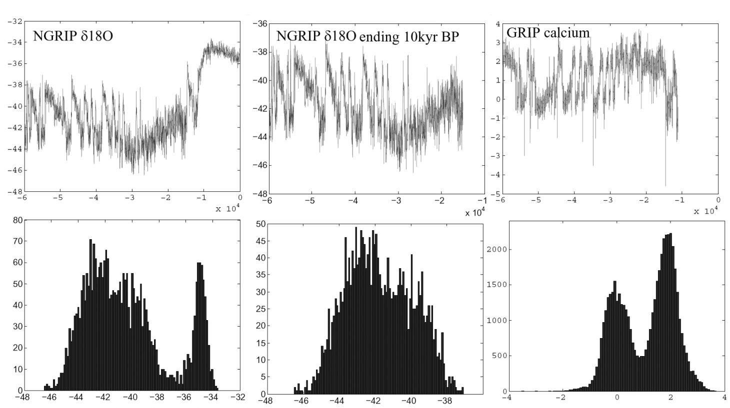


Figure 1: NGRIP $\delta^{18}O$ and GRIP Ca series and their histograms.

One can see that these two series, when compared on the same interval 60-10 kyr BP, have essentially different probability distributions. Whether the Langevin equation driven by $\alpha$-stable noise could be used to describe the NGRIP $\delta^{18}O$ record with resolution 20yr is an interesting topic of further research, but is beyond the scope of the present manuscript.

- *Finally, the method fundamentally relies on the 1-D nature of the state-space model, where the drift is parameterised as the gradient of a pseudo-potential. The difference between the sample autocorrelation obtained with this model, compared to data, is uncomfortable (figure 6 of KL). KL hypothesized that the memory at large lags, which they identified in the observations, may be due to the nonstationarity. Now that Livina et al. have identified a non-stationarity, would it be possible to test a model where this nonstationarity is effectively taken into accont and verify the shape of the autocorrelation function?*

  This is an interesting point for further research but we think it is outside the scope of the present paper which clearly focuses on identifying the number of states in the system.

- *As far as I recall Saltzman and Verbitzky did not adhere to the multiple-stable state paradigm, but rather interpreted the trajectory in the ice volume - CO2 phase space as the signature of a limit cycle.*

  This is correct, and we have removed reference to Saltzman and Verbitzky to avoid confusion. In fact, our approach can be extended to a second-order potential model which allows for limit-cycle behaviour as discussed by Kwasniok and Lohmann (2009b).

- *Equation (1) is inconsistent for purists of stochastic differential equations. $\eta$ should rather be written as the increment of a Wiener-process.*

  We have switched to the more mathematically accurate notation.

- *How do we know, after for example rejection of $L = 4$ due to negative a4, that a model with $L = 6$ would not provide a much better fit. Admittedly this makes intuitive sense, but a more rigorous justification would be welcome.*

  Indeed, our method is empirical; a more systematic approach may lie in minimising an information criterion which takes into account both goodness of fit and number of fit parameters.

- *How was the level of noise determined in the surrogate data?*

  In the surrogate data, the noise level was defined as $\sigma = 1.5$ in order to get a reasonable regime transition time scale. There is a note on this in section "Data".

- *Ditlevsen (1999) used Ca data rather than $\delta^{18}O$, the former offering a much better resolution. What justifies the present choice of $\delta^{18}O$?*

  We started studying $\delta^{18}O$ with best available resolution in public domain 20yr, to compare GRIP and NGRIP series. Later, we got access to the annual GRIP Ca series (courtesy of P. Ditlevsen), and we were pleased to detect a similar bifurcation in this series. The results on Ca are now also presented in the paper.

Please note that we adjusted the contour plots to map the results to the middles of the sliding windows instead of the ends, which is more natural way of plotting due to aggregation of histogram data within intervals.

We hope that the manuscript is suitable for publishing in the Climate of the Past.

Yours sincerely

V. Livina, F. Kwasniok, and T. M. Lenton