In this paper two paleoclimatic ice core records are analyzed. These, the water isotope, d18O, and the dust concentration records are analyzed for the period 59-27 kyr BP, which is the glacial period dominated by regular occurrences of Dansgaard-Oeschger events. There are two major points in the paper: Firstly, the data are modelled as a stochastic process using the Kramers-Moyal equation to investigate the importance of (discontinuous) jumps in the noise. Secondly, the two records are modelled as a two-dimensional joined process. It seems to me that the two points are only loosely related, and the authors could consider presenting them in two separate papers.

I enjoyed reading the paper and find it publishable. However, there are a few issues below calling for revisions before publication.

I have two major concerns regarding the two parts, and some minor points:

As to the first point, I have not seen the Kramers-Moyal (KM) equation applied to these data before, so this is a novel approach. The equation, for which the Taylor expansion of the conditional probability density function is taken to higher order than two, covers the
case where the noise term in the governing Langevin equation is not gaussian, but contains jumps. It is stated that in the case of Levy processes the Fokker-Planck equation does not apply. Actually, for the most relevant class of Levy processes, the alpha-stable Levy processes an extension of the Fokker-Planck equation based on the characteristic function of the alpha-stable process exists (see: Samorodnitsky and Taqqu (1994) 'Stable non-gaussian processes, Chapman and Hall, NY. or Ditlevsen, PRE, 60, 172-179). The challenge in applying the KM equation is the estimate of the higher order coefficients (eq. 6) for the data series: Since the higher order terms are (increasingly) dominated by the extremes in the increments the finite time series very quickly "dilutes". A main result (eq. 9 and figure 3) includes the sixth moment of the increments. I thus miss an analysis of uncertainties and reliability in these estimates. I find that this is essential for publishing this (nice!) result.

Another point which could be given a little more attention is the fact (as also correctly stated) that the strong time-asymmetry in the data (the sawtooth shape) cannot be captured by the model. How does this influence the relevance in including higher order terms (higher than second order) in eq. 5?

As to the second point, the major results are presented in figure 4. Obviously, when considering a one-dimensional record, the drift can always be seen as a result of a potential. This is not the case in two - and higher dimensions, where gradient drift is a non-generic case. I’m sure that the authors are aware of this, the drift is a two-dimensional flow field, as also shown in the small inserts in the subplots of figure 4. I find the construct of pseudo-one-dimensional potentials (V(x_1|x_2)) both confusing and useless. I suggest that the authors consider abandoning this all together (as well as the notion of a potential landscape). The interpretation in figure 4(c) of a double fold bifurcation is obscure, and -I believe- wrong. What I find interesting is the scatter plot in 4(a), which nicely explain the results in figure 2, namely that the stationary distribution for the dust is bimodal while it is unimodal for the d18O: This corresponds to the marginal distributions in 4(a) (projections onto the axis. The authors could consider analyzing a rotation (linear combination of the two variables) of the data along an axis connecting the two maxima (GS and GI) and a perpendicular direction. In this way one would obtain a “clean” two state dynamics and a “clean” one state perpendicular dynamics. First thing would be to check for independence. Just a suggestion.

Minor points:
Introduction: These data have been analyzed over many years and a lot is known. For a better overview and setting the present work in context, a representative presentation of work done over the years would be useful. There is a strong bias towards very recent publications.

Again: I’m not fan of the “potential landscape” metaphor.

Figure 1: I suggest to plot the dust record upside-down. This will visualize the strong dependence between the two records, and also make the saw tooth shape in the dust record much more apparent. Make the figure full text width, Ylabel: ln(dust) (no units), d18O (permil). Or normalized w.r.t. std. dev.

L90: A discussion of concentration vs flux of dust could be added

L91: Data is -> Data are

L96: This was pointed out by others previously: Rial and Saha (2011), Abrupt Climate Change: Mechanisms, Patterns, and Impacts, Geophysical Monograph Series 193, Mitsui and Crucifix, arXiv:1510.06290, Lohmann and Ditlevsen, 2018, Clim Past 14. (and probably others).
L110: I do not understand why

L125: Stationarity require certain properties of a(x) (such that the process does not drift to infinity. It is better to denote it “homogenous” or “autonomous”. (same comment on L138).

L126: Langevin equation is continuous  -> Langevin equation generate realizations which are continuous

L155: There is no Eq 4. (4a and 4b are hardly equations)

L185: The purely continuous process (gaussian process) diffuse proportionally to t^(1/2) not t. That is: (sigma(t)=sqrt(4Dt)). The most natural jump processes in this context are the alpha-stable processes, they do exhibit similar scaling relations with time (sigma(t)~t^(1/alpha)).

L205: Also referring back to Figure 1: Wouldn’t it be natural to rescale the data before doing the analysis.

L239: I do not understand the statement (which I believe is not correct): This indicates
that d18O exhibits faster dynamics than dust. Please explain.

L245: The exotic explanation for monostability through a complex noise structure seems a little out of context here: The reason why the d18O record has a single maximum in the PDF (with a shoulder) is the sawtooth shape of the DO events masking the obvious two state nature of the record.

Section 4.2.1 is obscure to me.

L273 and L288: 4 (d) <-> 4(c).

L456: What happened to the index _t in mu and sigma^2?

I hope my comments are useful.
