Answers: green
Comments by reviewer: violet

Comment on esd-2021-95

Tamás Bódai (Referee)

Referee comment on "Changes in stability and jumps in Dansgaard–Oeschger events: a data analysis aided by the Kramers–Moyal equation" by Leonardo Rydin Gorjão et al., Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2021-95-RC2, 2022

First of all, we would like to thank the referee for his thorough and detailed review of our submitted manuscript. His scrutiny from an outside perspective revealed that the manuscript is not as clear as it should be about some central points regarding its conceptual design. We believe that we can in this answer clarify our approach and clear up some misunderstandings. This review gives us the chance to substantially improve the manuscript's clarity with respect to the conceptual design, its explanatory power and its limitations.

The paper "Changes in stability and jumps in Dansgaard–Oeschger events: a data analysis aided by the Kramers–Moyal equation" analyses d18O and dust data from a Greenland ice core in order to gain further understanding of the famous Dansgaard-Oeschger (DO) events. They preprocess the time series with the aim of establishing a stationary stochastic process. They estimate Kramers-Moyal (KM) coefficients, which could possibly reveal jumps in the process, outside the framework of the Fokker-Planck equation, corresponding to what can be naively seen as regime transitions. They explore the added value of joint fitting of the d18O and dust data over treating them separately.

The referee correctly summarised the approach of our manuscript. There are two minor points we would like to clarify.

First, we do not claim that the nonzero 4th-order KM coefficient – which we interpret as evidence for a jump-like stochastic forcing on the d18o – fully explains the regime switches. We only say that the jumps in the stochastic forcing might play a role in the regime switches and should not be discarded in the analysis.

Second, the term 'fitting' usually describes a method where a model output is compared to the data. Then the model parameters are tuned such that the model would optimally approximate the data. The approach pursued in our study is different. We estimate the Kramers-Moyal coefficients directly from the data and no model-output to data comparison is required for this method; in particular, we do not minimise a distance or cost function. We deliberately refrain from presenting explicit stochastic model equations.

I'm not very convinced that the applied methodology is suitable. As far as i see, the authors do not test their null-hypothesis (H0) of a stationary process.

We fully agree with the referee that one cannot assume per se that the investigated time series are stationary. Also, the referee is correct in the sense that we did not provide a statistical test that supports the stationarity of the investigated time series. In the revised version of the manuscript we will employ a slightly different detrending of the data and provide tests that do support stationarity.

Since the referee's general suspicion towards the applicability of the chosen method is the key criticism in his review, we will in the following give a detailed explanation why we consider our approach meaningful.

We believe that a comment the referee made in the pdf attached to his report, is helpful to understand the exact point of criticism he raised with respect to the stationarity assumption:

In line 401, we write:

> *'This may be the atmospheric circulation as represented by the dust proxy, or another external driver whose signature might be encoded in the higher-order KM coefficients of the δ18O.'*

which was commented by the referee with the words:

> This sounds like there is no problem with the methodology if that's the case.

This comment led us to the interpretation that the referee does not question that the climatic process giving rise to DO events can be considered stationary over the investigated time period. Instead, we understand the referee's point as follows:

The observed data is a projection of a high dimensional complex process onto the state space spanned by $\delta^{18}O$ and dust which are assumed to represent Greenland temperatures and atmospheric large-scale circulation. The applied methodology now assumes that all other degrees of freedom (or all other variables, termed 'bath') can be subsumed in an effective force and a stochastic force (i.e., noise), and be described in a SDE approach. This subsumption certainly relies on the type of interaction of the observed variables with the bath variables and requires a separation of time scales.

We believe that the referee doubts whether the relevant dynamics that gives rise to the observed DO variability in the data at hand are fully captured in the projection onto the observed, low-dimensional subspace. We interpret his objection to our stationarity assumption in the sense that the referee advocates for the presence of unobserved or hidden variables that cannot be subsumed in the bath treatment. Such a potential coupling to hidden variables can also be interpreted as non-stationarities of the observed dynamics.

For sake of clarity: In the following we will refer to unobserved variables that cannot be described as a bath by denoting them as *hidden variables*.

If the requirements for the eff. force + noise description of the unobserved variables are not fulfilled and one still tries to impose this framework – as we do – then the hidden variables which in fact have much more explicit impact on the observed variables' dynamics, still contribute to the KM coefficients and in particular to higher order KM coefficients. One would then in most cases observe that the model retrieved from the data does not fully explain the dynamics of the observed variables.

At this stage, we can make three important remarks:

1. The observations are limited to the d18o-dust space, so all we can do is try to investigate these. We cannot include further variables in our analysis, simply because there is no data.

2. Given (1) and in line with finding the simplest starting argument, it is natural to make the attempt to treat the unobserved variables as a bath and then scrutinise the consistency of the obtained KM coefficients with the observations. There will typically be some characteristics of the dynamics which are reasonably well explained by this approach, and others which are not. This is exactly the case in our study.

   For example: In our 1D analysis of the $\delta^{18}O$ we emphasise the inconsistencies of the obtained KM coefficients with the data. It is these inconsistencies that motivate us to explore the next complicated approach of analysis which is the investigation of the coupled dust-d18o dynamics.

'At first sight, the monostability of the reconstructed $\delta^{18}O$ potential contradicts the apparent two regime nature of the time series. There are two possible explanations for this discrepancy: First, regime switching of monostable stochastic process can be achieved through complex noise structures (e.g., Lévy-like noise, generalised Fokker–Planck equations, or fractal motions) (Chechkin et al., 2003, 2004; Metzler and Klafter, 2004). Secondly, a similar effect can be obtained in a two-dimensional setting if the dynamics of one dynamical variable explicitly depends on the other, which would be impossible to judge from the one-dimensional analysis presented so far. Thus, within the limits of this analysis – that is assuming that the process is Markovian and stationary and that the system under study is fully represented by dust and δ 18 O (no coupling to further hidden variables) – the source of the regime switching must either be endowed by more complex noise processes or by the coupling between the dust and the δ18O systems.' (l.243)

3. If there would be a strong coupling to hidden variables, we would expect the data to show higher degrees of autocorrelation, since typically subsuming hidden variables mistakenly in a bath gives rise to a memory term. This essentially follows from the Mori-Zwanzig formalism.

In the 2D setting, we limit the KM analysis to the first and second-order coefficients, due to the scarcity of data. In our investigation we then focuses on the retrieved deterministic flow field and we do not claim to provide a comprehensive explanation for the dynamics of the coupled dust-δ^{18}O system. What we do find, however, is that this flow field does not, in itself, explain the fast transitions c2w and the slow w2c transitions – here, importantly, we point to the possibility that hidden variables may play a decisive role at various points in the manuscript.

'The effective vector field of motion does not indicate a clear path that the system would take in order to transition between stadial and interstadial states. This leaves open the possibility that transitions between stadial and interstadial states are mainly induced by noise as argued by, e.g., Ref. (Ditlevsen et al., 2007) (i.e. noise-induced tipping), facilitated by a shallow potential barrier close to the minima of the (effective) vector field.' (l.321)

(Which was commented by the referee as follows:

I'm really not convinced. I think the consideration of an important variable is missing.

'Evidently, our analysis is limited to solely two climate proxy variables and the stability of the one can only be assessed conditioned on the other, leaving aside potential coupling to further external factors.' (l.349)

'The non-vanishing fourth KM coefficient in the δ18O, which indicates forcing beyond typical Gaussian white noise, could point to an external trigger that directly acts on the Greenland temperatures.' (l.359)

'The results obtained in our analysis do not give a clear answer to the question for the exact mechanism that triggered DO events. In principle, the revealed double-fold bifurcation would allow for bifurcation-induced transitions and thus for a limit-cycle

behaviour. However, the records show that the system does not track the stable fixed point branches until the bifurcation points, but tends to transition earlier (not shown). Also, the structure of the δ 18 O drift is incompatible with a deterministic cyclic motion in the dust-δ 18 O plane. In fact, the specific structure of the double-fold bifurcation leaves room for a weak barrier between stadial and interstadial states in the vicinity of the bifurcation point, thus creating a 'channel'-like passage, through which the system passes.'

In short:

- We agree that a coupling to hidden variables which cannot be subsumed in a SDE representation can be understood as a type of non-stationarity with respect to the dynamics of the observed variables.
- We believe the referee understands our claim that the data is stationary in the sense that we categorically exclude any coupling to hidden variables. This was, however, not our intention. We further agree that the relation between the KM coefficients and potential not bath-like hidden variables is not explained sufficiently in the manuscript as is.
- In a revised manuscript we would therefore point out the fact that we are potentially missing important parts of the systems state space and elaborate on how this relates to our assumptions of stationarity and how hidden variables can influence the estimation of the KM coefficients if one imposes the analytical framework which is build on stationarity. Also, we will argue that based on the very small autocorrelation, it is reasonable to assume a negligible coupling to hidden variables as a first-order approximation.
- We will also explain more precisely the steps we take in our analysis right at the beginning of the manuscript. That is, we start with the simplest models, and then elaborate on what these models do explain and what they fail to explain. Then we move to the next complicated models and do the same. So far, no study that was concerned with modelling Dansgaard-Oeschger variability has claimed to explain the dynamics of these events in full detail.
- We would also like to remind the referee of the fact that only limited data from ice cores is available for these long time periods and with sufficient temporal resolution. So all we can do is to investigate the observed variables and there is little we can do about the hidden ones – at least if we aim to stick to a methodology which is to a high degree data-driven.

I would not think that a stationary process described by the KM equations is consistent with a hypothetical nonstationary process that could not be rejected.

In principle, it is certainly true that the stationary KM equation, which forms the basis of our investigation, is not a consistent framework to describe a non-stationary process. We have made an effort to rule out obvious reasons for non-stationarity in the manuscript:

'Excluding also the Last Glacial Maximum from the data, we restrict our analysis to the period 59–27 kyr b2k, which is characterised by a fairly stable background climate and persistent co-variability between dust and δ^{18}O. To compensate for the remaining influence of the background climate on the climate proxy records, we remove a linear

trend with respect to the global average surface temperature from both time series (see App. A for the details).'

For sake of clarity, in a revised of the manuscript, we will change the sentence:

'Excluding also the Last Glacial Maximum from the data, we restrict our analysis to the period 59–27 kyr b2k, which is characterised by a fairly stable background climate and persistent co-variability between dust and δ^{18}O.'

to

'Excluding also the Last Glacial Maximum from the data, we restrict our analysis to the period 59–27 kyr b2k, which is characterised by a fairly stable background climate, pronounced DO variability and persistent co-variability between dust and δ^{18}O.'

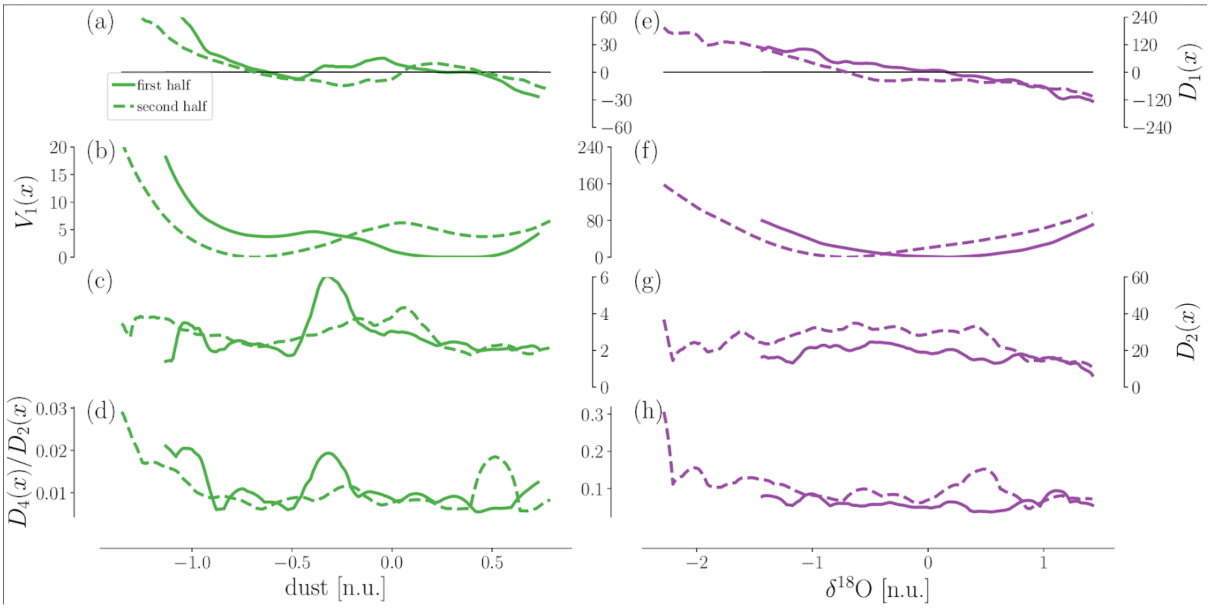
We include here a set of two unit root tests that indicate the data is stationary in the sense that there is no slow change in the process characteristics. These tests are the Augmented Dickey–Fuller test (ADF) and the Augmented Dickey–Fuller-GLS test (ADF-GLS). Both test for the possibility of a unit-root in the time series (null hypothesis). The alternative hypothesis is the time series does *not* have a unit root, i.e., it is stationary (in a broad sense).

The tests allow for different forms of trends behind the data. The ADF test allows for having solely a constant offset (no trend), solely a trend (no constant offset), a constant offset and a trend, and a constant offset, linear and quadratic trend. The ADF-GLS contains only a constant offset (no trend), or a constant and a trend.

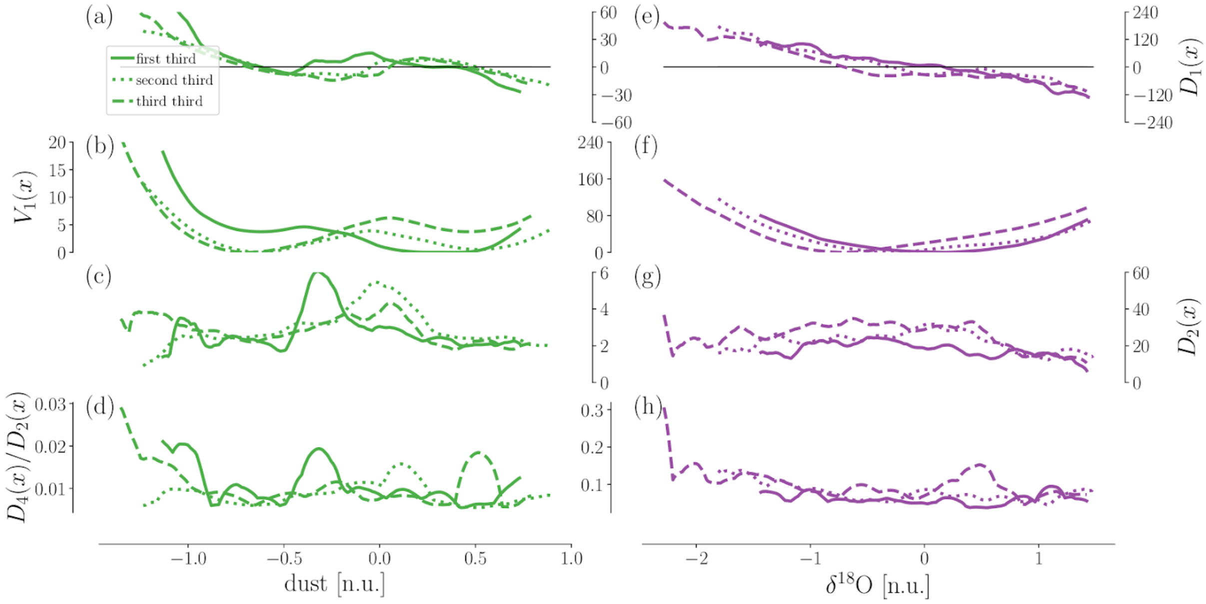
	ADF				ADF				ADF-GLS		ADF-GLS	
	Null Hypothesis: Non-stationarity Reject if statistics is *smaller* than critical values				**Null Hypothesis: Non-stationarity** Reject if statistics is *smaller* than critical values				**Null Hypothesis: Non-stationarity** Reject if statistics is *smaller* than critical values		**Null Hypothesis: Non-stationarity** Reject if statistics is *smaller* than critical values	
	dust				**δ^{18}O**				**dust**		**δ^{18}O**	
	no trend	no constant	constant and linear trend	constant, linear, and quadratic trends	no trend	no constant	constant and linear trend	constant, linear, and quadratic trends	constant	constant and linear trend	constant	constant and linear trend
	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]	statistics (*p*-value) [optimal lag]
critical value (p=0.05)	**-1.9410**	**-2.8620**	**-3.4112**	**-3.8333**	**-1.9410**	**-2.8620**	**-3.4112**	**-3.8333**	**-1.9470**	**2.8499**	**-1.9470**	**2.8499**
data	-4.9124 (1.566e-06) [6]	-5.0515 (1.753e-05) [6]	-5.3732 (4.082e-05) [5]	-5.481 (1.382e-04) [5]	-6.6017 (4.651e-10) [15]	-6.8034 (2.208e-09) [15]	-7.3196 (2.644e-09) [15]	-7.5415 (4.278e-09) [15]	-3.4422 (6.373e-04) [6]	-5.2217 (9.989e-06) [5]	-3.7747 (1.904e-04) [15]	-6.6558 (1.487e-08) [15]

All tests point to an absence of a unit root in our time series (results are valid also as p=0.01).

Secondly, we take the same KM analysis we performed and apply it in the first half and second half of the time series, to showcase the overall structure of the potentials/drifts remains unaltered.



We see that the overall shape of the drifts/potentials remains the same. Naturally, the zero-crossings of the drift change since in a one-dimensional analysis the coupling between the proxies cannot be considered in one-dimension. A similar recipe is taken again by dividing the time series into 3 subsequent thirds.



We note that this last case considers performing the KM analysis over a time series of ~2200 data points. Even in this regime of a very low number of data points, we observe the same double-well structure in the dust and single-well structure in the δ^{18}O.

After these considerations the only source of potential non-stationarity that is left is the variability of other climatic subsystems (e.g. the AMOC) which are potentially coupled to Greenland temperatures and atmospheric circulation investigated in this manuscript, that is the coupling to hidden variables.

Finally, it should be mentioned that several influential investigations of the same data have followed a similar reasoning, that is, they rely to some extent on the assumption that the data generating process can be described by autonomous model equations that comprise only the observed variable and no hidden variables. A small selection is:

> Ditlevsen, P. D. Observation of α-stable noise induced millennial climate changes from an ice-core record. Geophys. *Geophys. Res. Lett.* **26**, 1441–1444 (1999).
>
> Boers, N. *et al.* Inverse stochastic-dynamic models for high-resolution Greenland ice core records. *Earth Syst. Dyn.* **8**, 1171–1190 (2017).
>
> Kwasniok, F. Analysis and modelling of glacial climate transitions using simple dynamical systems. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **371**, (2013).
>
> Livina, V. N., Kwasniok, F. & Lenton, T. M. Potential analysis reveals changing number of climate states during the last 60 kyr. *Clim. Past Discuss.* **5**, 2223–2237 (2010).

Say, we have a nonstationary Orntsetin-Uhlenbeck process of dx/dt = -a*x + B(t) + c*xi(t), (OU) where xi is white noise, and B(t) = b*sin(sin(2*t)+t) is some regular nonstationarity. It mimics some regime behaviour with sudden and regular transitions. We can easily see that the pdf of x is bimodal. If we didn't know the underlying process generating eq., and perhaps we somehow overlooked the regularity of the transitions, we might think the underlying model is:

dx/dt = F*x + c*xi(t), (H0)

where F = -V(x), V(x) being a double-well potential function. If we are in the small noise limit, we know that the pdf takes the shape of V(x), so, we could estimate V that way. Furthermore, we can estimate the noise strength ‘c’ in some standard way too.

Now the question is if it matters at all that we have an H0 other than the true process OU. It would not matter if in any appreciable way the processes perform the same, i.e., when, loosely speaking, they are consistent or approximately equivalent. For example, we can derive the probability distribution of residence times from H0, and perform a statistical test if our residence time data is consistent with that, or we can reject H0.We want to perform a so-called crucial experiment (experimentum crucis).

We start our investigation by estimating the KM coefficients up to 4^{th} order of the isolated 1D time series for both, δ^{18}O and dust. In the case of the dust, this does immediately imply a Langevin type model, since D_4 is negligible. In the case of δ^{18}O, we do not propose a full model since there are several SDE models which are consistent with the retrieved KM coefficients.

As proposed here by the referee, we then test the consistency of the results (that is of structure of the drift and diffusion and the D_4 in case of the δ^{18}O) with respect to the data. We explicitly emphasise that the data contradict the recovered KM coefficients in the 1D case. This motivates us to consider the coupling between δ^{18}O and dust in the next more complicated setup. In 2D, we do not test a specific model, since the correspondence between KM coefficients and model coefficients is not trivial.

> *‘At first sight, the monostability of the reconstructed* δ^{18}O *potential contradicts the apparent two regime nature of the time series. There are two possible explanations for this discrepancy: First, regime switching of monostable stochastic process can be achieved through complex noise structures (e.g., Lévy-like noise, generalised Fokker–Planck equations, or fractal motions) (Chechkin et al., 2003, 2004; Metzler and Klafter, 2004). Secondly, a similar effect can be obtained in a two-dimensional setting if the dynamics of one dynamical variable explicitly depends on the other, which would be impossible to judge from the one-dimensional analysis presented so far. Thus, within the limits of this analysis – that is assuming that the process is Markovian and*

stationary and that the system under study is fully represented by dust and δ^{18}O (no coupling to further hidden variables) – the source of the regime switching must either be endowed by more complex noise processes or by the coupling between the dust and the δ^{18}O systems.'

Considering H0 of the authors, another likely feature based on which H0 can be rejected is the saw tooth asymmetry, in particular, that the cold to warm, c2w, transitions are much more rapid than the warm to cold, w2c, ones.

If we consider the setting of a standard and stationary Langevin process as

$dx(t) = - a(x)dt + bdW(t)$,

Where $a(x)$ is mean-reverting, we indeed have a process that is one-dimensional and time symmetric. We can easily break the time symmetry by introducing a discontinuous element, as for example

$dx(t) = - a(x)dt + bdW(t) + cdJ(t)$

where $J(t)$ is a Poisson process with a jump rate $\lambda>0$. In this simplest of formulations, if there is at least 1 jump from the Poisson jump process, the process becomes time-asymmetric. (From an applied point of view, there are some requirements regarding the relation of the amplitudes of $a(x)$, b, and c, and the smoothness and differentiability of $a(x)$.) Thus, breaking time-symmetry in a one-dimensional setting is not impossible, on the contrary, a single discontinuous trajectory does the job.

We do not explicitly mention this or any other explicit stochastic process, but instead show that there must exist discontinuous trajectories in our time series (Fig. 2, lower panels with the $D_4(x)/D_2(x)$ ratio and Fig. 3, depicting the *Q*-ratio). This offers one answer to why we can have time asymmetry even in a one-dimensional setting. As correctly stated by the reviewer, it nevertheless does not exclude a potential presence of non-stationarities in the data. We note that in two dimensions time asymmetry is easily created by a state-dependent drift term $a(x,y)$.

As a last point, we agree there are other methods to estimate the various parameters in the system, as mentioned in regards to estimating $F(x)$ or c. In our manuscript we have focussed solely on the shape of $F(x)$, and in a very agnostic manner, avoid discussing the exact functional values of $F(x)$ of amplitude c.

Looking at the dust time series of Fig. 1 with much naivity wrt. physics, but with some experience about dynamical systems, I would think that c2w is an attractor crisis, whereas w2c is a noise induced tipping, and there is some slowly drifting control parameter, i.e., a nonstationarity when we exclude that parameter from our state variables.

The authors also make a reference to attractor crisis, in terms of a saddle-node bifurcation, but in some other context. It is the context of slices of a 2D potential function. This does not sound correct.

Following from the feedback of the other reviewer, even in a setting where in one dimension one has a double well and in the other a single well one might find a spurious double fold bifurcation in a KM analysis under the right rotation of the basis vectors. As we now explain, while it is not possible to reach a definite answer on the entire physical mechanism leading to the observed bimodality from our approach, we can derive some very interesting conclusions, as we now discuss in detail. Please see the answer to the other reviewer.

The paper is very well written in a way, but it doesn't make for a very pleasant reading journeying through flawed results, starting with the single variable approach, and then — at least as i suspect — even the 2 variable approach.

I attach the pdf of the manuscript with comments saved as annotations. Hopefully the authors find it useful in some way.

Note: I always review non-anonymously, and never make recommendation for or against publication. The recommendation that i make is only to circumvent the rigidity of the submission system, and therefore please consider it void.

Tamas Bodai

Please also note the supplement to this comment:

https://esd.copernicus.org/preprints/esd-2021-95/esd-2021-95-RC2-supplement.pdf

We thank the reviewer for his valuable comments. We agree that, particularly using proxy data in the context of paleoclimatology, much of the outcome is subjected to limitations and imperfections. Still, we believe our proposed revisions will take this into account thoroughly and provide important new conclusions e.g. about how the two variables, dust and $\delta^{18}O$, are coupled. As discussed in the answer to the other reviewer, the widespread assumption that the bistability is rooted in the temperature, which then drives the dust variable as a dependent variable, seems unlikely from our results.

The particular aspect of including both 1D and 2D approaches follows from the fact that the 1D approach is a natural starting point for our investigation also considering other recent studies on date-driven model inference from Greenland ice core time series, and provides the reason why the 2D analysis is necessary, therefore motivating the second step of analysis.

Specific Comments

---

l.12 (3) the $\delta^{18}O$ record is discontinuous in nature, and mathematically requires an interpretation beyond the classical Langevin equation.

A discrete time sampling cannot result in a continuous series. Also, is this new result?

We agree with the reviewer that a discretised time series cannot be continuous, but this is not the point here. What we express in the paper is not a relation of the continuity or discontinuity of the discretised continuous time series, we refer solely to stochastic processes (not their discretised versions or measured time series). These, mathematically, can be continuous or discontinuous. The result is, to the extent of our knowledge, new. It arcs back to previous results that suggest, e.g., other potential discontinuous stochastic models to describe the proxies (Lévy processes), yet here we include direct estimations of textbook measures to show that the data is indeed discontinuous (mathematically).

l.32 The most prominent example of past abrupt climate shifts are the Dansgaard–Oeschger events

maybe one of the most

Thank you for this comment. We will replace the statement by:

One of the most prominent examples of past abrupt climate shifts are the Dansgaard–Oeschger events

l.34 a series of sudden warming events that dominated Greenland temperatures throughout the last glacial cycle, ~~e.g., Refs.~~ (Johnsen et al., 1992; Dansgaard et al., 1993; North Greenland Ice Core Projects members, 2004).

e.g., Refs was stroke through by the referee. We will change to:

[...] *a series of sudden warming events that dominated Greenland temperatures throughout the last glacial cycle, (e.g. Johnsen et al., 1992; Dansgaard et al., 1993; North Greenland Ice Core Projects members, 2004).*

l.57 The key concept is to regard the paleo-climate record as the realisation of a Markovian and stationary stochastic process (Kondrashov et al., 2005, 2015) which can be described in terms of a stochastic differential equation.

What if the sudden warming is a bifurcation

Here, we refer to our detailed answer to the referee's general comment.

l.63 The Kramers–Moyal equation generalises the Fokker–Planck description of stochastic processes, including explicitly the presence of discontinuous elements.

Is this equivalent with a nonstationary framework? I suppose not.

There is no difference here, the Fokker–Planck equation is a limit case of the Kramers–Moyal equation for vanishing higher-order Kramers–Moyal coefficients. The formulation of a non-stationary Fokker–Planck process is identical to that of a Kramers–Moyal process, it simply involves defining the Kramers–Moyal coefficients with a temporal dependency. See Risken (1996) chapter 4 for a general derivation of the non-stationary Kramers–Moyal equation and how to constrain it to a Fokker–Planck equation aided by Pawula's theorem. There are no limitations or different impacts of non-stationarity in either case.

l.67 In Sec. 2 we introduce the paleo-climatic proxies under examination and the detrending method used to ensure that the data is approximately stationary.

Is this correct methodology? Why would the detrended process not be nonstationary? I.e., is stationarity really "ensured"?

The detrending removes the anyways small non-stationarity that stems from the slowly varying background climate. However, it does not remove potential non-stationarities of the sort we believe the referee has in mind, namely those which are due to couplings to hidden variables. A series of tests have been included in our earlier reply above.

l.72 This is consequently discussed in Sec. 4.2, where we uncover the conditioned potential landscapes of the joint proxy process.

conditioned on what?

Conditioned on the respective other variable. However, in view of the comment by the other referee, we will abandon the notion of conditional potential anyways and rephrase this sentence accordingly.

l.109 To assess whether the data is Markovian, we analyse the auto-covariance function of the increments of the detrended data.

Nonstationarity overlooked this way.

Please see our detailed answer to the referee's general comment.

l.121 A prominent example for a stochastic process is given by the stationary Langevin equation

I thought in a Langevin eq. you have the differential on the left. This looks more like Ito or Stratanovic, and you should actually specify which one.

The formulation is given in Itô.

l.124 If the properties of the dynamics do not change over time, i.e. a(x) and b(x) do not depend on time, these processes are called stationary.

This is what i would contest.

Please see our detailed answer to the referee's general comment.

l.125 While the Langevin equation is continuous in time, stochastic processes can in principle have discontinuous features, such as sudden jumps.

But for what value of dt? var[dB] goes to 0 as dt does.

Discontinuity in our case is not related to a sampling rate. Please see the explanation below where we now explain in detail what is considered continuous/discontinuous in our formulation.

l.126 An easy way to incorporate discontinuities is to include in Eq. (1) an elementary Lévy process L(t), modulated with an amplitude h(x) (Applebaum, 2011)

Isn't it a bit of an interpretational issue? var[dL] does not exist. But then what' in the limit of 0 for dt?

The inexistence of the variance has no relation with the continuity of a stochastic process. The variance of a Poisson process always exists (all moments exist) and this is truly a discontinuous process. Please see the explanation below on Lindeberg's continuity condition.

l.135 If a single particle's motion is governed by the Langevin equation, its probability density function p(x, t|x ′ , t ′) evolves according to the Fokker–Planck equation,

Does the FP not govern the unconditional density rather?

Both forms exist. See Risken (1996) eq. 4.16 for a conditional density formulation (with the general Kramers–Moyal operator), or eq. 4.52 and 4.53 for solely the conditional density formulation of the Fokker–Planck equation.

l.140 So, this should tell whether (1) is Ito or Stratanovic. Please spell it out. My reference is Risken's book for this.

The formulation is given in Itô.

l.145 Giving up the condition of continuity, the temporal evolution of the conditional probability density follows the Kramers–Moyal equation

Can you please indicate here how discontinuity is allowed?

Consider the Lindeberg's continuity condition *C(t)* for a markovian process, which states that a trajectory *x* is continuous if

$$
\begin{aligned}
C(t) \triangleq C(x,t,\delta) &= \lim_{\tau\to 0^+} \frac{Prob[\ |\Delta x(t)| > \delta \quad |_{x(t)=x}]}{\tau} \\
&= \lim_{\tau\to 0^+} \frac{\int_{|\Delta x(t)|=|x'-x|>\delta} p(x', t+\tau|x,t)\, dx'}{\tau} = 0
\end{aligned}
$$

For all $\delta>0$, all x and t, with $\Delta x(t)=x(t+\tau)-x(t)$ and $p(x',t+\tau|x,t)$ the conditional prob. density. Take a stationary distribution for a Langevin-like process, characterised by a drift $D^{(1)}$ and a diffusion $D^{(2)}$, and no higher-order KM terms ($D^{(4)}=D^{(3)}=0$, and Pawula's theorem), and insert it into the Lindeberg's continuity condition

$$
\begin{aligned}
C(x,t,\delta) = \lim_{\tau\to 0^+} \frac{1}{2\tau\sqrt{\pi D^{(2)}(x,t)\tau}} \Bigg\{ \int_{-\infty}^{-\delta+x} \exp\left(-\frac{(x'-x-D^{(1)}(x,t)\tau)^2}{4D^{(2)}(x,t)\tau}\right) dx' \\
+ \int_{x+\delta}^{\infty} \exp\left(-\frac{(x'-x-D^{(1)}(x,t)\tau)^2}{4D^{(2)}(x,t)\tau}\right) dx' \Bigg\}
\end{aligned}
$$

Which we separate into two integrals:

$$
\equiv \lim_{\tau\to 0^+} \frac{1}{2\tau\sqrt{\pi D^{(2)}(x,t)\tau}} (I + II)
$$

With I as

$$
\begin{aligned}
I &= \int_{-\infty}^{-\delta+x} \exp\left(-\frac{(x' - x - D^{(1)}(x,t)\tau)^2}{4D^{(2)}(x,t)\tau}\right) dx' \\
&= \sqrt{4D^{(2)}(x,t)\tau} \int_{-\infty}^{(-\delta - D^{(1)}(x,t)\tau)/\sqrt{4D^{(2)}(x,t)\tau}} \exp(-u^2)\, du \\
&= \sqrt{4D^{(2)}(x,t)\tau}\, \frac{\sqrt{\pi}}{2}\, erfc\left[\frac{\delta + D^{(1)}(x,t)\tau}{4D^{(2)}(x,t)\tau}\right]
\end{aligned}
$$

Take an approximation of the error function expanded in τ

$$
\begin{aligned}
I = 2D^{(2)}(x,t)\tau \frac{\exp\left\{-(\delta + D^{(1)}(x,t)\tau)^2/4D^{(2)}(x,t)\tau\right\}}{\delta + D^{(1)}(x,t)\tau} \times \\
\left\{1 - \frac{2D^{(2)}(x,t)\tau}{(\delta + D^{(1)}(x,t)\tau)^2} + \frac{12(D^{(2)}(x,t)\tau)^2}{(\delta + D^{(1)}(x,t)\tau)^4} + \cdots\right\}
\end{aligned}
$$

In the limit of $\tau \to 0^+$, I=0. *Mutatis mutandis* for II. This proves the continuity of a Langevin-like process. The central assumption here lies on the fact that $D^{(1)}$ and $D^{(2)}$ are sufficiently well behaved and smooth.

We now consider the case wherein there are higher-order moments and we consider directly the conditional moments and a limit $dt \to 0$, such that

$$
M^{(m)}(x) = \lim_{dt\to 0} \tfrac{1}{dt} \langle |x(t+dt) - x(t)|^m |_{x(t)=x} \rangle
$$

From here we show that

$$
\lim_{dt\to 0} \frac{Prob[|\Delta x| > \delta\ |_{x(t)=x}]}{dt} \leq \frac{M^{(m)}(x)}{\delta^m}
$$

Which, importantly, is *not* zero. Follow our definition of the Kramers–Moyal and conditional moments as given in the paper to arrive at

$$
\begin{aligned}
\langle |u - x|^m \rangle|_{x(t)=x} &= \int_{-\infty}^{\infty} du\ |u-x|^m p(u, t+dt|x,t) \\
= \int_{-\infty}^{x-\delta} du\ |u-x|^m p(u,t+dt|x,t) &+ \int_{x-\delta}^{x+\delta} du\ |u-x|^m p(u,t+dt|x,t) \\
+ \int_{x+\delta}^{\infty} du\ |u-x|^m p(u,t+dt|x,t)
\end{aligned}
$$

From which we construct

$$
\langle |u - x|^m \rangle|_{x(t)=x} \geq \int_{|u-x|>\delta} du\ |u-x|^m p(u,t+dt|x,t)
$$

by ignoring the second term in the 3 integrals above. Use now that

$$|u - x|^m > \ \delta^m$$

Take the limit as above to obtain

$$\lim_{dt \to 0} \frac{1}{dt} \langle |u - x|^m \rangle|_{x(t)=x} \geq \delta^m \lim_{dt \to 0} \frac{1}{dt} \int_{|u-x|>\delta} du \ p(u, t + dt|x, t)$$

Which is the Lindeberg condition above. Previously we saw that Langevin-like equations result in an equality with *zero*. The existence of any higher-order KM coefficients dilutes this to yield a non-zero value, proving the discontinuity. This inequality can be made into a strict equality in certain cases, like the aforementioned stochastic process with Poissonian jumps. In the case of Poissonian jumps, we know that the discontinuity is simply given by their jump rate λ.

[These derivations are taken from MRR Tabar (2019) *Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems*]

l.153 However, for numerous of these stochastic processes, the KM coefficients can be related to the properties of the stochastic process in the spirit of Eq. (4).

Provide a reference please.

In a revised version of the manuscript we will cite Risken (1996), Tabar (2019), and Anvari *et al.* (2016).

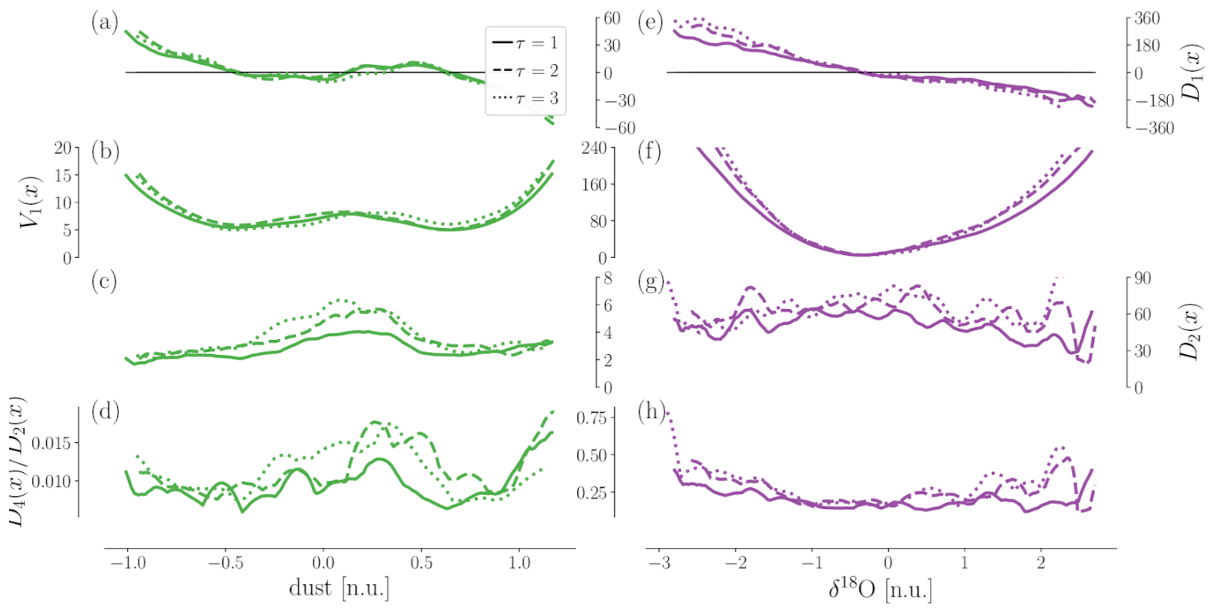
l.168 To retrieve the KM coefficients D m (x) from a single realisation of a stochastic process, i.e. a single time series, we evaluate the transition probability densities in the limit of a vanishing time step $\tau \to 0$, which numerically corresponds to considering the shortest increment Δt in the data ($\tau \to \Delta t$).

Is it so? I suspect that it might be but might be not. It can be that you are trying to do your estimation of a process, which requires a higher resolution of data.

Essentially, you might be unable to perform the estimation (even if the modeling assumption was completely perfect). What you want to see perhaps is that as you reduce the time resolution, your estimate seems to converge. I.e., you might need to deliberately coarsen your resolution.

Below we include a reploting of Fig. 2 in the manuscript wherein we coarsen the resolution τ from the original 1 to 2 and 3 (i.e., we consider only half and a third of the total amount of data). The overall shapes remain unaffected. Bandwidth selection of the Nadaraya–Watson estimator still follows Silverman's optimal bandwidth selection, using an Epanechnikov kernel. The total amount of data-points considered are: $\tau=1$, 6399 data-points; $\tau=2$, 3199 data-points; $\tau=3$, 2133 data-points.

We note that following our observation from the auto-correlation of the increments of the time series, seen in Appendix B, Fig. B1, we can safely claim we are above the Einstein–Markov length and thus these estimates are not affected by microscopic noise correlation and subsequently taking coarser resolutions yield identical results (just as shown), which are only affected by having fewer data-points from which to draw the estimates.



I think it is not a good procedure to perform the estimation and then generate ample synthetic data by the assumed model and check if estimates at a certain resolution are biased.

We assume the referee inserted the ‘not’ inadvertently and hence we interpret this comment as if the referee suggested to apply the above procedure. In this case our answer would be as follows: To the best of our understanding, this is not something we can perform in our evaluation: First, we can naturally estimate all statistical moments of the time series, but to generate synthetic data we have to make an assumption on the underlying process. For instance, the Tabar Q ratio presented in Figure 3 shows the presence of discontinuities or jumps in the process. We refrain from formulating a specific model for these jumps on the basis of the limited data we have. Our paper goes to great length to avoid precising a model – that we believe is one of the strengths of the paper. Thus, unless we constrain ourselves to a model, we cannot do this.

If the ‘not’ was not a typo, then we do not fully understand this comment as we do not simulate any synthetic data.

It is possible that at the biased estimates the biases are indicated low, but at the true "unbiased" values estimates are very biased at that time resolution. It might sound circular, but it isn't, i believe.

l.178 Thus, values of the ratio D 4 (x)/D 2 (x) close to zero imply continuous sample paths with no jumps in the data.

The estimate will be a finite nonzero number. How do i know if this is small, close to zero, or not? What should i compare it with?

The term is proportional to the sampling rate τ as $M^2 \sim \tau$ and $M^4 \sim \tau^2$ (M the conditional moments) if we are considering a Langevin process. Thus, the smaller the sampling rate, the smaller the ratio. For the case of jumps, the relation $M^4 \sim \tau^2$ is no longer valid.

l.183 This assessment can be refined by regarding the Lehnertz–Tabar Q-ratio (Lehnertz et al., 2018), which takes advantage of the fact that continuous and discontinuous systems ‘scale’ in a different fashion. While a purely continuous stochastic process diffuses proportionally to time t (or possibly a power of time t^β in anomalous diffusions (Einstein, 1905; von Smoluchowski,1906; Havlin and Ben-Avraham, 1987)), discontinuous processes can cover large distances in short times, i.e. jump, which causes them to exhibit no scaling relations with time t.

I don't know what this means.

We will include an appendix in the revised manuscript detailing what we mean with discontinuity, how is the Kramers–Moyal formulation a candidate to describe discontinuous processes, and how to understand the scaling relations in order to understand the Q-ratio. We shall as well reformulate this passage and subsequently point to the appendices that best explain what is meant with scaling.

l.205 For various applications where the fluctuations are not comparable in size, i.e. where the diffusion elements are not of similar scale, one can draw a clearer picture of the motion of the two-dimensional system by referring to an effective vector field

Reference please.

In fact, we did not adopt this approach from other studies. Thus we will replace the above statement by:

Given the different levels of diffusion along the two dimensions, we introduce here an effective vector field by rescaling the drift components by the value of the corresponding diffusion in each direction.

l.211 Similarly to the one-dimensional case, one can obtain potential landscapes as integrals over the two drifts:

Reference please. How is the potential defined over a multidimensional phase space? I think there is only one potential function, not one per variable! See p. 133-134 of Risken, or https://iopscience.iop.org/article/10.1088/1361-6544/ab86cc e.g. eq. (3)

But mind that you can only say that you found a potential from obs data if your modelling assumptions are correct.

In a two dimensional state space one can consider the dynamics along a single dimension while hypothetically freezing the motion along the other dimension. This leads to a 1D setting, where the typical notion of a potential applies.

However, we decided to abandon the notion of the conditioned potentials and will therefore reformulate statements and the equations accordingly.

l.228 We find the second KM coefficient to be fairly constant (Fig. 2 (c)) and the ratio between fourth and second KM coefficients to be negligible (Fig. 2 (d)), which suggests that a Langevin process with additive noise is a viable description of the isolated dust dynamics.

I cannot judge whether this is large or small.

Please see our answer to the referee’s comment to line 178.

l.233 Note that the model equations employed here are by construction symmetric with respect to time, therefore, as it is, the model cannot reproduce the temporal asymmetry that is visually suggested in the dust record.

This should raise concern about the modeling assumption.

In fact, the model equations only become symmetric wrt. time after estimating the diffusion to be constant over space and the D4 coefficient to be negligible. They are not symmetric by construction as explained in our detailed answer to the referee's general comment. We will therefore rephrase:

'Note that with the second KM coefficient being constant and the fourth KM coefficient being negligible, the obtained model equations are symmetric with respect to time. Therefore, as is, the model cannot reproduce the temporal asymmetry that is visually suggested in the dust record.'

l.236 Most prominently, the drift has only a single stable fixed point (zero-crossing of the drift), or equivalently, the potential function exhibits only a single well.

The drift implies a fixed point, maybe. Terminology.

Yes, indeed. We will replace the statement by:

'Most prominently, the drift has only a single zero-crossing, or equivalently the potential function exhibits only a single well.'

l.240 Moreover, we find that the fourth KM coefficient D 4 (x) for the δ 18 O is of the same magnitude as the second KM coefficient D 2 (x) (Fig. 2 (h)).

I see that D4 is compared to D2 actually. But what does comparability, a ratio of 1, mean? Still, is this much in some sense?

Please see our answer to the referee's comment on line 178.

l.239 Given the high correlation between the dust and the δ18O records, the differences in the reconstructed potentials and the ratio between fourth and second KM coefficient are remarkable.

Yes, but in the same time we can just plot the histograms of dust and d18O and we will see a bi- and unimodal distribution, no? IF the 1D models were correct, and the noise was really small, then the histograms give you the respective potential functions.

It is true that already the PDF of dust and $\delta^{18}O$ suggest a qualitative difference between the two time series. We will integrate the PDFs of both records in the figure 4 of the revised manuscript. However, the correspondence between a records histogram and the underlying potential can easily be corrupted by multiplicative gaussian noise, let alone more complex noise. So the qualitative difference in the potentials can only be evidenced in this more comprehensive type of analysis.

Maybe thought the very correlation bw. dust and d18O prompts that the 1D models cannot be good assumptions.

Sorry, we are not sure we understand this comment.

One needs to somehow test the hypothesis of the model. Checking Markovianity in the appendix is probably very insufficient.

Please see our response to the referee's general comment above.

l.243 At first sight, the monostability of the reconstructed δ 18 O potential contradicts the apparent two regime nature of the time series.

Is the histogram bimodal? Bimodality would imply regime behaviour (of whatever origin), but regime behaviour does not imply bimodality of the hist'.

We fully agree with the referee. The histogram of the $\delta^{18}O$ is in fact monomodal. However, a two-regime nature of the record can indeed be evidenced by eye. In the manuscript we introduced the term two-regime nature maybe with a lack of explanation. We would therefore add a paragraph on this already in the data section.

l.258 In Fig. 3, we clearly see a constant relation of $Q(x, \tau)$ with respect to τ for the $\delta^{18}O$ record, suggesting that this stochastic process includes jumps.

Or that your assumption of stationarity is crucially wrong. We want certainty. There is little value in evidence alone that can imply two very different things. We want to know which one is true.

Please see our reply to the referee's general comment.

l.274 The reconstructed conditional potential $V_{1,0}(x_1|x_2)$ of the dust is displayed in Fig. 4 (d). As a conditioned potential, it can be read by taking vertical 'slices' of the potential.

You haven't defined that. And so we don't see that the 1D potentials are really slices of the one multivariate pot'.

As mentioned previously, we will abandon the notion of conditioned potentials, and hence reformulate the above statement accordingly.

l.290 However, the position of the minimum δ 18 O* appears to be determined by the dust in a continuous manner, with high rate of change for intermediate dust values whilst no change for more extreme dust values.

What does this refer to?

We meant to say that for intermediate values of the dust a small change in the dust – seen as a control parameter – causes a fairly strong change in the position of the minimum of the conditioned $\delta^{18}O$ potential (high rate of change).

As we abandon the notion of conditioned potentials, we will rephrase the above statement.

l.297 These findings are consistent with the observed regime switching of both records, which we struggled to reconcile with the results obtained from the one-dimensional analysis.

Perhaps there is an amount of hindsight in this but this is exactly how time series would look like when we have a 2D double well potential with two axes of symmetry.

Do not define the variables by the axes of symmetry, and you will see coordinated transitions. Furthermore, if you tip your coordinate system only slightly, closing small angles with the axes of symmetry, then you have the chance that despite the transitions, the marginal distribution of one of the variables will be still unimodal.

However, such a system would not feature asymmetry in that you have a more sudden transition in one direction than the other.

We fully agree with the referee's comment. Since this point is hardly a criticism, we do not intend to change the manuscript in response to this. We do point out the possibility to obtain two-regime-like, yet unimodally distributed time series in one variable, if this variable is suitable coupled to another one.

'*At first sight, the monostability of the reconstructed δ 18 O potential contradicts the apparent two regime nature of the time series. There are two possible explanations for this discrepancy: First, regime switching of monostable stochastic process can be achieved through complex noise structures (e.g., Lévy-like noise, generalised Fokker–Planck equations, or fractal motions) (Chechkin et al., 2003, 2004; Metzler and Klafter, 2004). Secondly, a similar effect can be obtained in a two-dimensional setting if the dynamics of one dynamical variable explicitly depends on the other, which would be impossible to judge from the one-dimensional analysis presented so far.'* (l.243)

l.323 This leaves open the possibility that transitions between stadial and interstadial states are mainly induced by noise as argued by, e.g., Ref. (Ditlevsen et al., 2007) (i.e. noise-induced tipping), facilitated by a shallow potential barrier close to the minima of the (effective) vector field.

I'm really not convinced. I think the consideration of an important variable is missing. It's worth considering this recent paper:

https://link.springer.com/article/10.1007/s00382-020-05476-z

The sawtooth feature is there, for one thing. But my feeling is that the rapid transition is rather due to crisis, the disappearance of a regime, as a result of nonstationary dynamics (nonstationary in a reasonable modeling framework).

Please see our detailed reply to the referee's general comment.

l.358 In particular, we have shown that the δ 18 O record cannot be treated as a time-continuous process.

If the fast transition is attractor crisis, then it is a continuous process. You can only make this statement if the modelling assumption was water tight. But it isn't. So, strictly speaking, you have not shown what you say.

In the light of what has been said before, the statement should be made more precise:

'In particular, we have shown that the isolated δ18O record cannot be treated as a time-continuous process in a one-dimensional SDE setting.'

l.364 In principle, the revealed double-fold bifurcation would allow for bifurcation-induced transitions and thus for a limit-cycle behaviour.

What you found is not what you say. I think the methdology of (13) is problematic.

We are not sure if we fully understand this comment. We will not use (13) in the revised manuscript and instead explain the presence of the double-fold bifurcation in terms of the nullclines of dust in the $\delta^{18}O$, dust state space.

l.372 We conclude therefore that – based on our results – the DO transitions are to large degree induced by noise, acting on the background of a double-fold dynamics governing the dust, for which the δ18O acts as control parameter.

I'm not convinced at all, sorry. How could the faster dynamics control the slower like this?!

We agree that this statement should be attenuated by saying:

'In the two dimensional subspace spanned by $\delta^{18}O$ and dust, assuming that no couplings to hidden variables substantially influence the dynamics, our results suggest that the DO transitions are to large degree induced by noise, acting on the background of a double-fold dynamics governing the dust, for which the $\delta^{18}O$ acts as control parameter.'

l.373 These findings do not contradict previous studies that proposed limit-cycle models to explain the δ18O record, since the cyclic motion was not expected to happen in a state space comprised of Greenland temperatures and atmospheric large scale circulation (Kwasniok, 2013; Lohmann and Ditlevsen, 2019).

I think your finding is the result of your methodology. It is your methodology that should/could be contradicted. Consider the possibility that the cited papers do contradict your methodology.

I'm not sure how, but some statistical test might be able to reject your null-model. Perhaps the saw-tooth asymmetry feature can be a basis of such a formal, precise test.

Or maybe another test is whether the residence time data is consistent with the predictions of the fitted model. Of course, i mean a formal hypothesis test again.

There is really no value in further considering models that can be rejected by data outright.

A question is whether negative results should be published.

Many studies experiment with simple models in order to investigate Dansgaard–Oeschger variability. Some of those models – as a double well potential plus noise (Lohmann, 2018; Kwasniok, 2013; Livina, 2010) – can *a priori* be seen to not capture essential features of the NGRIP $\delta^{18}O$ time series. Nevertheless assessing their performances, and finally deciding they would not explain the phenomena under study is a valuable contribution.

The difference in our case is that due to the ambiguity of the 4^{th}-order KM coefficient we do not propose a full model in order to simulate the process. We acknowledge that the referee does not believe that DO variability can be explained in the dust-$\delta^{18}O$ state space and we will emphasise the possibility of missing a crucial dimension of the dynamics in our analysis in the revised version of the manuscript.

I think it depends on how nontrivial they are. So, it should be considered by the authors that looking at Fig. 1 a stationary model is trivially wrong.

We have already explained why even a stationary 1D model is not trivially wrong.

l.416 Our analysis considered only a two-dimensional projection of the very high-dimensional dynamics and can therefore not be expected to deliver all details of the triggering mechanism.

Regarding the merit of a study, we can consider the question: can attractor crisis be modelled by a stochastic process that features some large jumps?

We do not fully understand the question. An attractor crisis as a potential trigger of c2w transitions would in first place need a higher dimensional state space. However, as already mentioned, we aimed to carry out a data-driven study, but no additional data is available for the investigated time period at the required resolution.

l.417 Neither the ocean-focused self-sustained oscillation hypothesis, nor the idea that the atmosphere acts as a trigger, can be ruled out based on our findings.

Isn't this a problem?! Wouldn't it be a problem that your paper did not reduce uncertainty about which one it is?!

Of course, we had hoped that our investigation would more clearly point to either the one or the other direction. However, the results are as they are and still contain valuable information.

l.422 Finally, our study underlines the need for higher resolution data, as the scarcity of data points is a limiting factor for the quality of non-parametric estimate of the KM coefficients.

It would have been good to see in an appendix the dependence of estimates on time resolution.

We assume the referee refers to his comment made in the general comments section:

Essentially, you might be unable to perform the estimation (even if the modeling assumption was completely perfect). What you want to see perhaps is that as you reduce the time resolution, your estimate seems to converge. I.e., you might need to deliberately coarsen your resolution.

Please see the ‘General Comments’ section for our answer.