Point-by-point response to reviewers and summary of manuscript changes

Jakob Harteg, Nico Wunderling, Jonathan F. Donges

April 29, 2026

Cover letter

We thank the editors and both reviewers for their careful comments. We respond point by point below. A high-level summary of changes is in Section 1; the uploaded `latexdiff` marks changes line by line.

1 High-level summary of changes

Main revisions at a glance (referee attribution where the prompt came from; other edits are consistency or presentation).

- Clarified the mathematical structure of the model: differential–algebraic formulation (ice volume v prognostic; CO_2 and T diagnosed each step), discrete update order, and why combining the diagnostic equations does not contradict the equation for $\dot{v}$. *Reviewer 1.*
- Corrected a **sign error** in the published CO_2 diagnostic so it matches Talento & Ganopolski (2021) and the code.
- Clarified orbital forcing: Talento–Ganopolski composite index (not summer-solstice insolation alone), preprocessing of precession/obliquity inputs, and consistent anomaly notation $(f - \bar{f})$ in text and figures. *Reviewer 1.*
- Added a finite-time Lyapunov exponent (FTLE) diagnostic (Mitsui & Crucifix, 2016) on the same stochastic ensemble as $\mathrm{RAR}(t)$, as a *comparison* metric. *Reviewer 1.*
- Updated the combined RAR/FTLE figure (Figure 3 in the current numbering) to overlay several α_{RAR} values; removed the separate supplementary figure that only showed RAR sensitivity to tolerance.
- Renamed the "single-event perturbation" ensemble to "pulse-perturbation" ensemble; kept the name Reference Adherence Ratio but write $\mathrm{RAR}(t)$ throughout to stress time dependence. *Reviewer 2.*
- Noise sweep figure: reduced from five to four σ levels (dropped the weakest) so the panel reads more clearly; the main stochastic analysis still uses $\sigma = 10^{-3}$.
- Stochastic ensemble for $\mathrm{RAR}(t)$/FTLE: reduced from 20,000 to 1,000 realisations, because the larger sample was redundant for the reported diagnostics (outcomes unchanged in practice). The pulse-perturbation experiment still uses its original ensemble size.
- Clarified notation: separate tolerances α_{RAR} (stochastic band for $\mathrm{RAR}(t)$) and α_{RT} (return-time criterion); ensemble members $\tilde{v}_k(t)$ vs. reference $v(t)$ in the stochastic experiment, and $\tilde{v}$ in the pulse-perturbation return-time definition; return time written as $T_R(t_p)$ with explicit perturbation time t_p.
- Regenerated or adjusted all figures for consistent style, axes, and labelling.

- References were brought in line with the revised text (added citations for new discussion, removed duplicate Berger/PAGES entry for the same paper, and updated published metadata where needed).

2 Reviewer 1 (Takahito Mitsui)

1. Inconsistency in the model equations.
When $\frac{dv}{dt} \geq 0$, Eq. (3) is given as

$$CO_2 = c_1 T - c_2 v + c_4\,.$$

Eliminating T using Eq. (4),

$$T = d_1 v + d_2 \log\left(\frac{CO_2}{278}\right)$$

yields

$$CO_2 = (c_1 d_1 - c_2)\, v + c_1 d_2 \log\left(\frac{CO_2}{278}\right) + c_4$$

This expression implies

$$v = \frac{CO_2 - c_1 d_2 \log\left(\frac{CO_2}{278}\right) - c_4}{c_1 d_1 - c_2}$$

which contradicts the dynamical equation for v given in Eq. (1). A similar issue also arises for the case $\frac{dv}{dt} < 0$.

Eqs. (3)–(4) are diagnostic closures, not prognostic equations: combined, they give an implicit algebraic relation among CO_2, v, and (for $\dot{v} < 0$) $\dot{v}$, as expected in a differential–algebraic formulation. Section 2 now states that *only* $v(t)$ is prognostic, with $CO_2(t)$ and $T(t)$ diagnosed each step and documented update order (backward-difference $\dot{v}$ for CO_2, then advance v).

We corrected a **sign error** in the published Eq. (3); the manuscript now uses

$$\mathrm{CO_2} = c_1 T + c_2 v + c_3 \min\left(\frac{\mathrm{d}v}{\mathrm{d}t}, 0\right) + c_4 + \mathrm{Anth}_{\mathrm{CO_2}},$$

consistent with Talento & Ganopolski (2021, their Eq. 7) and our implementation. With this correction, the reviewer's elimination still yields an implicit diagnostic constraint—not a contradiction with Eq. (1), because CO_2 is not an independent state variable.

2. Relation to previous studies on stability and resilience.
This reviewer views the work by Harteg et al. as valuable in that it explicitly places glacial–interglacial cycles within the modern framework of Earth system resilience theory. On the other hand, the resilience or (in)stability of glacial–interglacial trajectories has been discussed in several earlier studies.

Finite-time Lyapunov exponents have been used as indicators of transient instability in De Saedeleer et al. (2013), Mitsui and Aihara (2014), and Mitsui and Crucifix (2016). In these studies, the finite-time Lyapunov exponent increases prior to the temporal separation of unperturbed and perturbed trajectories (e.g., Fig. 8 in Mitsui and Aihara, 2014; Fig. 6 in Mitsui and Crucifix, 2016). This behavior appears slightly different from that of the RAR. For instance, the RAR remains nearly constant during trajectory separation near MIS 3 in Fig. 3 of the present manuscript. One might ask whether log(RAR) could serve as a more sensitive indicator.

In addition, so-called potential analysis has revealed temporal changes in attractors or basin structures across glacial–interglacial cycles (Livina et al. 2011). Dakos et al. (2008) discussed conventional early-warning signals prior to deglaciations, although these may not be applicable to strongly non-equilibrium systems. A clearer comparison with these existing approaches would strengthen the manuscript.

We thank the reviewer for highlighting this literature. We have strengthened the Introduction and Discussion and added an FTLE comparison in the main stochastic diagnostics figure. Our points follow the order of the comment:

- **MIS 3 / near-flat** RAR(t)**:** For the main α_{RAR}, many stochastic trajectories are already outside the tolerance band during that part of the record, so the adherence *fraction* can sit close to zero and vary only weakly while trajectories separate—including near MIS 3 in Fig. 3. The revised main figure overlays several choices, $\alpha_{\mathrm{RAR}} \in \{0.001, 0.0025, 0.005\}$, so how strongly return behaviour depends on the threshold is visible directly (replacing a separate supplementary sweep).
- log(RAR)**:** We examined log(RAR) as suggested. It increases sensitivity mainly when RAR $\ll 1$ but does not change the qualitative phase interpretation. We therefore retain RAR(t) as the primary quantity and do not add a separate figure to this response document.
- **FTLE:** We compute the windowed FTLE from Mitsui & Crucifix (2016), with a small regulariser ϵ when separations are near zero and with $\tau = 10$ kyr (a deglaciation-scale horizon; not the memory timescale τ in the ice equation). In line with our first response, the same stochastic realisations show intervals of convergence toward the reference around interglacial onset and divergence in glacial phases, while a feature near MIS 3 can be comparatively subtle next to stronger convergence–divergence episodes elsewhere. Panel (c) of the combined stochastic diagnostics figure in the revised manuscript (Figure 3) plots this alongside RAR(t): FTLE reflects local growth of infinitesimal separations over a fixed window, whereas RAR(t) is the fraction of realisations within a finite tolerance band of the reference.
- **Potential analysis and early-warning signals (Livina et al., Dakos et al.):** In the Discussion (*Relationship to existing research*), we contrast those approaches—which infer stability structure from palaeoclimate time series—with our emphasis on path-wise behaviour under explicit finite perturbations relative to the same forced reference trajectory.

> **3. Importance of the proposed resilience measures.**
> In previous studies, finite-time Lyapunov exponents have been used to explain the existence of periods during which the simulated ice-age trajectory is vulnerable to perturbations. Although this article emphasizes the necessity of path-wise resilience concepts, it remains unclear how the proposed RAR and the return time can be used in practice. For example, the reviewer wonders how these measures can be useful for assessing the safety of long-term storage and disposal of nuclear waste, which was one of the original motivations of the work by Talento and Ganopolski (2021), whose model is adopted in Section 2. If this question is beyond the scope of the present study, it can be safely ignored.

We agree that the specific choice of indicator is not the central contribution. The aim is to introduce and demonstrate a path-resilience perspective for a non-autonomous glacial–interglacial trajectory, for which many classical (autonomous, attractor-based) resilience indicators are hard to interpret. RAR(t) and return time are therefore used as simple illustrative diagnostics: in practice, they highlight phases of comparatively higher or lower robustness to finite perturbations (rapid re-convergence vs. sustained separation). We do not claim direct applicability to nuclear-waste safety assessment; that application lies outside the scope of the present study, and we therefore do not discuss it in the manuscript.

> Section 3.1.3: This study assumes a unique deterministic trajectory in the absence of perturbations. However, previous work has demonstrated the possible coexistence of multiple trajectories (e.g., De Saedeleer et al. 2013). Moreover, in more complex climate models (including CLIMBER-2), deterministic trajectories can be chaotic and may depend on initial conditions, although they may

> remain largely synchronized with astronomical forcing. In such cases, defining a reference trajectory vref (t) becomes non-trivial. This issue does not necessarily undermine the usefulness of the RAR; however, it would be helpful if the authors discussed how the proposed framework could be applied in such situations.

We agree that a single "reference trajectory" is not always uniquely defined. Different parameter choices (as in Talento & Ganopolski (2021)) naturally give different deterministic trajectories, so in our analysis $v_{\mathrm{ref}}(t)$ is always meant for a fixed parameter set. A harder case is when, even with fixed parameters, the system can have multiple coexisting trajectories or be chaotic and depend on initial conditions. In that situation one could define $v_{\mathrm{ref}}(t)$ for a chosen initial condition, or define it from an unperturbed initial-condition ensemble (e.g. using the mean/median and the ensemble spread as a tolerance). We added a short paragraph about this in the Discussion.

> Line 121: "In the model, orbital forcing f represents maximum summer insolation at 65°N." Conventionally, maximum summer insolation at 65°N refers to the mean daily insolation at the summer solstice, for which the relative contributions of obliquity and climatic precession are defined a priori. In Eq. (5), however, their relative weights are tuned. This distinction should be clarified.

We agree with the reviewer's distinction. In our manuscript $f(t)$ is not the conventional "mean daily insolation at the summer solstice at 65°N". It is the orbital forcing index used in Talento & Ganopolski (2021), defined as a weighted linear combination of climatic precession and obliquity (Eq. (5)), with γ calibrated in that model. We have revised the text at the orbital-forcing passage the reviewer cited (former Line 121) to state this explicitly and to refer to $f(t)$ as a "65°N summer-insolation proxy / orbital forcing index" to avoid confusion.

> Line 124: It would be helpful to describe how $\mathrm{pre}(t)$ and $\mathrm{obl}(t)$ are preprocessed. The raw amplitudes of climatic precession and obliquity (expressed in radians or degrees) differ substantially.

We state that $\mathrm{pre}(t)$ and $\mathrm{obl}(t)$ are the series supplied in `smx_p.mat` / `smx_o.mat` (insolation-like scale), used without extra normalisation; mean subtraction of obliquity enters as in Eq. (5); $\bar{f}$ is taken over the 2000 kyr forcing window.

> In the caption of Fig. 1, f is the orbital forcing and f' is its anomaly. However, f appears to be anomaly in Eq. (5). Also $\mathrm{mean}(f)$ and $\bar{f}$ coexist.

We harmonised notation: figures and text use $f(t) - \bar{f}$ for the anomaly entering the ice equation, with consistent caption wording.

> Line 148: The term "a steady state" is not appropriate for a non-autonomous system. A pullback attractor, or simply a trajectory after removal of transients, would be more accurate.

We revised the wording accordingly.

> Line 173: $\mathrm{d}W_t$ is not the Wiener process; W_t denotes the Wiener process.

We corrected the text to refer to W_t as the Wiener process and $\mathrm{d}W_t$ as its increment.

> Figure 2: "Mean period" →"Period"?

The axis is now period (kyr); the caption states that the black curve is the ensemble-mean spectrum across runs.

3 Reviewer 2 (Anonymous)

> The modelled ice volume anomaly is normalised to zero at preindustrial level and unity at the last glacial maximum. Then the parameters of tolerance, etc. are set to proportional values to serve this specific dataset. I think if the authors like to achieve generality with the novel measures they introduce, they need to scale them with parameters of an arbitrary time series that may be considered in future. This generality would help adoption of the measures in the research community. What the authors demonstrate in Fig. 2 is the choice of the parameter value based on the fluctuations of the data. Can this be derived in general terms based on the properties of the input data?

We thank the reviewer for this observation. The measures introduced here are intended as a proof-of-concept demonstration within the specific model framework of Talento & Ganopolski (2021), where ice volume is normalised to $[0, 1]$. We agree that generalising $\mathrm{RAR}(t)$ and return time to arbitrarily scaled variables would require expressing the tolerance parameters relative to some property of the input data (e.g. its range or variance), and we have flagged this explicitly as a direction for future work in the revised manuscript. What we demonstrate with Fig. 2 is not a closed-form derivation from general properties of an arbitrary time series; it is a heuristic noise-amplitude sweep in this model, included as a consistency check across σ.

> I disagree with the statement (lines 210–213) that what rapidly happens during deglaciations has anything to do with resilience. Resilience means stability of a system state, when the system trajectory reliably fluctuates inside sufficiently deep potential well. Deglaciation is not that kind of behaviour; rather it is a transition from one state to another. If you drop one or two or three stones from a tower, they all will do something very similar, but this will not be resilience. I understand that the authors would like to interpret resilience as "the capacity of the ice-climate-carbon system to remain close to its characteristic largescale path in face of external shocks" (lines 35–36). But if the planet suddenly becomes hot, massive ice melting is alike a falling stone, and there are no less or more "resilient" alternatives to that.

We agree that, under the classical attractor/potential-well notion of resilience, deglaciations are transitions between regimes and should not be interpreted as "stability of a state". Our use of the term instead follows a *path-resilience* framing for a forced, non-autonomous system, where the reference is a time-dependent trajectory under the same external forcing. In this sense, the relevant diagnostic is whether perturbed realisations stay close to, or return toward, that reference path. This could also be phrased as the resilience or stability of the underlying pullback attractor. We made this clearer in the manuscript and, additionally, revised the wording around lines 210–213 by replacing

> "More indicative of genuine resilience are the increases in RAR ... The convergence ... points to enhanced dynamical stability intrinsic to the system ... "

with

> "More indicative of path resilience in our sense are the increases in $\mathrm{RAR}(t)$ that occur during the deglaciation phases—periods of rapid ice loss—when the constraint is not yet active and the system is free to evolve. Here the convergence of ensemble members toward the reference trajectory indicates an attracting segment of the forced trajectory, rather than stability of a fixed climate state."

> I think what the resilience function offers in Fig. 3 is a clear marking of deglaciation episodes (except N13) under the specific stochastic perturbation, which is interesting. The model was perturbed with a minor stochastic term, thus producing sufficiently similar trajectories driven by the planetary forcing—in particular, reproducing massive deglaciation when the coupled variables create conditions for that. In the context of the definition quoted above (lines 35–36), this stochastic perturbation is indeed resilient for a large number of simulations. Does it mean that the Earth system is resilient during the deglaciation?

In this paper, resilience is defined operationally as the tendency of perturbed trajectories to remain close to or return toward a reference trajectory under identical forcing. Within this framework, Fig. 3 indicates that deglaciations act as convergence phases for small perturbations in our model experiment. We emphasise, however, that this does not imply resilience of a stable state in the classical potential-well sense, but rather enhanced path-wise stability of the forced trajectory under the prescribed perturbations. We do not read this operational diagnostic as a universal claim about "the Earth system" in the abstract; model limitations (including the $v \geq 0$ bound) remain important caveats.

> In the plot with the resilience function (Fig. 3), I can understand the meaning of most of the maxima of the function except the one at 420–400 kyr. There, perturbed and reference datasets (upper panel) are in discord, and it is difficult to see these as staying within a small tolerance. Is there an explanation for that?

RAR(t) indicates that during 420–400 kyr about 60% of ensemble members lie within the tolerance band around the reference trajectory. In panel (a), those trajectories are visually obscured by the reference line, while the remaining members show a wider spread, which can make the ensemble appear more discordant than the tolerance-based count implies.

> In the definition of the return time (Eq. 9), in case of very long return times, the value of infimum will be the length of the considered time series. In the text, you mention that you run simulations 1000 kyr longer than 800-length series (line 230), and this explains why in Fig. 4, lower panel, at 100 kyr the return time in blue shaded area reaches 300 or so in value, while the remaining series is much shorter than that. To avoid confusion, I think the caption should have a note that the simulations were 1800 -long rather than 800 long.

Thank you for pointing this out. We agree that the caption of Fig. 4 should state that the simulations ran for an additional 1000 kyr beyond the plotted analysis interval; we amended the caption accordingly (800 kyr analysis window; 1800 kyr total length per run).

> Expression "single-event ensemble" is confusing, as it suggests that an ensemble may consist of a single event (which is not what the authors mean). It is better to say "single-event-perturbed ensemble"

We agree that the reviewer-suggested alternative is more precise, but it is also relatively long. Instead we use **pulse-perturbation ensemble** throughout for brevity and clarity.

> What the authors call "Reference Adherence Ratio" is actually not a ratio (a value) but a function of time. Therefore, it is better to name it Reference Adherence Ratio Function (RARF).

We agree that RAR(t) is time-dependent in our analysis. While the term "ratio" can sometimes suggest a single scalar value, it is also commonly used for quantities evaluated at each time step and analysed as a function of time. To avoid any ambiguity, we clarified this explicitly in the manuscript by consistently writing RAR(t), instead of just RAR, and referring to it as the time-dependent RAR. We therefore prefer to retain the name Reference Adherence Ratio (RAR) rather than introduce the alternative term "RARF".

> In Figure 2, panels b with spectra—in the 4th from top spectrum, there is an appearing state at about 80 kyr, which later disappears. Can the authors explain this effect?

This is a small but interesting feature. Since the model tends to go to full ice collapse once melting starts (through the M_v term in Eq. (1)), longer glacial cycles become less common as the noise level increases (higher σ). Around $\sigma = 10^{-2}$, the $\sim$100 kyr peak is already starting to break down. While we have not analysed this feature in detail, we hypothesise that the weak bump around $\sim$80 kyr is likely just the $\sim$100 kyr peak shifting/broadening as the cycle lengths shorten under stronger perturbations. The revised noise figure shows four representative σ levels with clarified labelling (Section 1).

> Line 195, it would be helpful to mention that both ensemble and reference are modelled datasets (the reference is not a reconstruction but rather a deterministic model trajectory).

We agree and replaced

"To quantify how strongly the ensemble trajectories cluster around the reference path"

with

"To quantify how strongly the simulated ensemble trajectories cluster around the simulated reference path"

> The preprint by Lucarini and Chekroun has already been published: `https://www.nature.com/articles/s42254-023-00650-8`

Thank you, we updated the reference.

> LaTeX typesetting of text after formulas in display mode: the authors apparently leave a blank line, which leads to indentation of the next line, although it continues the sentence started by the equation. This happens in lines 102, 111, 115, 126, 172, 199. This issue will disappear if the authors remove the blank lines after the display formulas preceding these lines.

Thank you, we fixed this.

> Line 24: I am not sure that "revitalise" is the right term for a planetary resilience. "Reinforce" may be a better term.

We thank the reviewer for the suggestion. We amended the manuscript: "Such knowledge is crucial for developing strategies to regenerate and reinforce Earth system resilience" (using "reinforce" rather than "revitalise").

> Line 117, paleo-reconstruction of what? (state the variable).

Thank you for pointing this out. The text now states that Figure 1 compares the simulated ice-volume anomaly with the palaeo-reconstructed global ice-volume anomaly derived from the sea-level stack of Spratt & Lisiecki (2016), along with the imposed orbital forcing and frequency spectra (Section 2). The Figure 1 caption reads: "Model validation against a palaeo-reconstruction of global ice-volume anomaly over the past 800 kyr. . . " with panel **(b)** specifying comparison to the global ice-volume anomaly reconstructed from that sea-level stack.

> Line 175, "We apply this"—what this?

We updated line 175 to state "We apply this stochastic differential equation framework to the ice-volume equation, which is the only differential equation in our model."

> In Fig. 2, vertical axes are missing

Thank you. We added the missing y-axis labels to Fig. 2.

> Line 316, I think system state cannot be "novel"; I would call them "previously unsampled". Or simply "new"

We replaced "novel" by "new".

> Line 370 , "RAR sensitivity to the tolerance value"

We originally indicated we would retitle supplementary material on RAR sensitivity to the tolerance value. In the revised manuscript, we instead **integrated** multi-threshold RAR(t) curves into the main RAR/FTLE figure (Figure 3), so sensitivity to α_{RAR} is visible without a separate appendix-style section; the standalone supplementary tolerance-only discussion is removed.