We complement our response to reviewer #3 with a detailed point-by-point answer (marked in blue).

Review of Talento and Ganopolsky paper on “Evolution of the climate in the next million years: A reduced-complexity model for glacial cycles and impact of anthropogenic CO2 emissions”.
May 28, 2021

1 Conclusion : Rejection

Glacial-interglacial cycles being very slow processes require a “proficient” and a “master” model for long-term temporal prediction. Predictions are carried by phenomenological models often represented as low-order dynamical systems. Low-order or reduced-order dynamical models (often represented by a set of coupled differential equations) are tractable and insightful by emphasizing the most important dynamical features of the complex behavior of a given system, such as the (paleo)climate system, in our case. In the latter, important information about that complex behavior is lost because of the use of tractable equations leading to under-defined parameters in the model representing the underlying phenomena. In addition, these models are data-driven, which make their calibration/the estimate of the model parameters and the forecast/prediction sensitive to the errors and uncertainty in the observational data. Therefore, using them to reproduce the current time and/or forecast the future, based on the past one, is highly dependent on the care and level of accuracy in calibrating and validating the model under consideration and specification of uncertainties. Combining Physical representations with probabilistic estimations is a very strong adequate way to forecast long term climate, especially in Paleoclimate (Crucifix and Rougier 2009). This requires the following steps:

1. Using a low-rder model to capture the very long term (millennial) of climate, under physical constraints. For instance the the three-dimensional stochastic system of Saltzman andMaasch (1991).
The authors here, designed/formulated three equations, using knowledge of the behavior of ice on millennial scales. Many assumptions has been advanced without any strong argumentation/reason/justification.

Apart from the fact that the reviewer recommends us to use the Saltzman and Maasch (1991) model instead of our own (which is lacking “*any strong argumentation/reason/justification*”), the meaning of this sentence with several typos and missing words is not clear to us. Not mentioning the fact that the model is designed not for millennial but for orbital and longer time scales. As far as the choice of the modelling approach is concerned, we must state that while we have a great respect to the works Burry Saltzman and his colleagues made during the 90’s, at present this and similar models are only of historical interest. Saltzman’s model is based on the assumption that glacial cycles represent self-sustained oscillations in the Earth system, and to produce such oscillations a cubic power term has been introduced in the equation for carbon dioxide without any justification. Since the 90’s our understanding of Earth system dynamics advanced significantly but no one was able to simulate self-sustained oscillations in the Earth system with realistic models and no one ever discovered this cubic power term. Needless to say that Saltzman and Maasch (1991) and similar models cannot be used for simulation of the impact of anthropogenic CO_2 emission on climate. To the contrary, our simple model is to a large extend based on the results of our own comprehensive Earth system model CLIMBER-2 which is able to simulate successfully not only the latest glacial cycles (Ganopolski and Brovkin, 2017) but also all Quaternary glacial cycles (Willeit et al., 2019).

2. Treating the estimate of the model parameters and the forecast probabilistically. One way of doing that, as in Crucifix and Rougier (2009) to assess the next glacial inception, is by inferring the different parameters within a Bayesian framework that allows for (1) parametric uncertainty and (2) for the limitations of the model, by using Sequential Monte Carlo technique (‘particle filter’).

As we show in our paper, the problem is not the treatment of model parameters “probabilistically” but the fact that past climate data provides no sufficient constrains on model parameters suitable for future simulations. We found that the majority of model versions which successfully simulate past glacial cycles have unrealistic relationship between the critical insolation threshold and CO_2 concentration and thus cannot be applied for the future simulations. In such a situation, any attempts to attach "objective" probabilities do different model realisations is nothing more than quackery.

3. Verify the accuracy and validating the model statistically and by checking the reproduction of physical phenomena. Different physical assumptions may lead to dynamical systems with dynamical properties that are similar enough to produce a convincing visual fit on palaeoclimate data [61]. challenge is, therefore, to operate a model selection on more stringent criteria than just fitting some standard time series.

Which *physical phenomena* reviewer means here we do not know, but we fully agree with the next reviewer's sentence: "*Different physical assumptions may lead to dynamical systems with dynamical properties that are similar enough to produce a convincing visual fit on palaeoclimate Data*". Indeed, during the recent decades different workers proposed a number of completely different mathematical manipulations which transform some combinations of the Earth's orbital parameters into curves with variability patterns more or less similar to the glacial cycles of the late Quaternary. Moreover even the use of a "*Bayesian framework*" did not help M. Crucific (Crucifix, 2012) to distinguish between the right models simulating glacial cycles as nonlinear response to orbital forcing, and the wrong models where glacial cycles originate from self-sustained oscillations. This is, of course, not surprising – the correct model cannot be derived solely from paleodata.

The approach which we employ in this study, and which represents a further development of the method used in Archer and Ganopolski (2005), is based on using a combination of paleoclimate data and the results of physically-based Earth system models. We believe, this is the only feasible alternative to the use of complex Earth system models, which are by far too computationally expensive for this task. In the "*critical comment which should be absolutely addressed*" #8 the reviewer wrote "*validating a work using results from another simulation [Ganopolski et al., 2016] does not seem accurate to me*". What "*accurate*" means in this context we do not understand. And, of course, we did not validate one model by another one. Instead, we use the results of the physically-based and well-tested Earth system model CLIMBER-2 to constrain parameters of a simple semi-empirical model which cannot be constrained by paleodata. We do not believe that there is an alternative to our approach.

In the rest of the review, the reviewer repeats time and time again that the right approach is the approach described in Crucifix and Rougier 2009 (hereafter CR09, the reviewer cited this paper nine times) and that our approach is absolutely unjustified (the reviewer used expressing containing "(un)justified" and "justify" more than 20 times!). We were glad to learn from Crucifix's comment that the reviwer#3 is not Michel Crucifix. We have great respect to Michel Crucifix whom one of the authors (AG) knows for 20 years since the time when Michel Crucifix was PhD student and AG was a member of his PhD committee. However, the methodology described in CR09 is absolutely inappropriate for our purposes. The main reason is that, although CR09 manuscript is entitled "On the use of simple dynamical systems for climate predictions", the authors of this manuscript used the term "climate predictions" with a meaning different from the one usually used. CR09 is about modelling of future glacial cycles without any anthropogenic influence. Of course, "climate prediction" usually means modelling of climate response to the anthropogenic perturbation. This is obviously the central goal of our study. The model which has been used in CR09 is not suitable for this task and the methodology described in CR09 is of no use for development and testing of such a model.

Besides, the Bayesian approach is not the only one possible or correct. Parameter estimation can be approached either through the frequentist or Bayesian point of view. In the frequentist framework point-estimates of unknown parameters are obtained and it is not possible to assign probabilities to the parameter values. It is assumed that there are enough measurements to derive useful information on the parameters. In the Bayesian approach the unknown parameters are treated as random variables and the measurements are complemented with information about a prior belief about the parameter values. Results may vary depending on which prior is selected. We opted for the frequentist approach and, therefore, there is no probability associated with the parameter estimation.

The necessary steps (1 to 3 above) required to assessing the future glacial inception under different levels of carbon dioxide emissions, have been inadequately followed and their related approaches incorrectly applied by the authors. The work by Talento and Ganopolsky does not reflect any aspect of the correct modeling approach towards a probabilistic forecast of climate. This work stated that, what is needed is a "quantitative probabilistic assessments"
as a must to assess on a very long term of carbon dioxide emissions on changes in temperature. As stated, this can be useful under the present challenges of climate change requesting carbon dioxide storage, which then requires
an adequate assessment of storage system under changes in the future environment due to human activity".
I do agree. However, this has not been done here. This is what the authors tried (wanted?) to accomplish, but failed unfortunately. This work has no provided any forecast neither probabilistic forecast of the climate. What has been done is a scenario simulation given a low-oder model (and even that, has has been inadequately assessed). They compared to a control simulation of future temperatures, where the anthropogenic emissions are null, a set of predicted simulations under low, medium, and high level of emissions. As carbon dioxide influences the coupled system temperature and ice on a long scale, they proposed a simple model, to be able to simulate a very long term of climate. In addition, the statistical modeling part, is applied incorrectly and many chosen assumption are unjustified. The model selection procedure (which model, among different alternatives, explains the observations best) has not been carried correctly either. No future forecast, or prediction (even under scenarios), especially when using observations, can be carried out non probabilistically. And when dealing with time series, it is even more critical to attach more attention to (1) more adequate statistical approaches for long term and multiple steps ahead forecast and to (2) adequate model validation and selection, where the predictive ability of the model must be verified given the length/characteristics of the observations (here, paleorecords). I explicated all these aspects in the document, where, I tried, despite the low level of the manuscript, to advise a way to correct the statistical modeling part, improve the paper, and follow a better predictive approach. The authors must chose one of the two research axes proposed below.
This paper cannot be published as it is and must be rejected. This work is not mature enough for publication. It needs a profound revision and rework. Concepts are being mixed and the goal itself is unclear to the authors.
The framework and the selected statistical modeling/validation approaches are weakly justified and poorly and/or incorrectly applied and most importantly the methodology is inadequate as it does not account for any source of uncertainty.
In a clear way and a more direct construction of the paper flow please, in a new version of the paper, chose one of these working axes:
1. Reconsider the whole work by implementing a probabilistic forecast approach, refer to Crucifix and Rougier (2009) and Crucifix(2012). Here, the inference should imply confronting a model with observations. "This inference process may take the form of a calibration procedure (update our knowledge on parameters on the basis of observations) or a model selection procedure (which model, among different alternatives, explains the observations best)" (Crucifix 2012).
2. Correct and adapt this work to reflect the framework of scenario simulation using a pre-constrained simple model. One way to make it publishable is to reformulate the goals and to position the work in literature related to
scenario based for decision making and not as a new probabilistic model for ice ages forecast (at all!). This part will require repositioning of the work in a more adequate framework, adapting the corresponding review of literature, choosing a correct approach for calibration, designing experiments under constraints for the optimization process (during the calibration process, to sample values of the parameters with appropriate sets of combinations under constraints) and fixing the vocabulary and giving a more adequate justification for all modeling choices.

As stated before, the CR09 approach is not applicable in our case.

We agree that the goal of the manuscript could be made clearer and that there is a need of position this work better within the existing literature. Following this reviewer and the other reviewers' comments, the word "forecast" will be substituted by "possible future scenario".

Regarding the reviewer' statement*: "As stated, this can be useful under the present challenges of climate change requesting **carbon dioxide storage**":* Our model is designed for supporting development of the "nuclear waste storage" not for "*carbon storage*". This model in principle cannot be used for carbon storage because does not allow negative CO_2 emission.

2 Main Comments

The authors formulated their predictive model as consisting of a system of three coupled non-linear differential equations, representing physical mechanisms relevant for the evolution of the temperature using a coupled Ice Sheets – Carbon cycle System in timescales longer than thousands of years, for different selected emission scenario. Many constraints have been introduced, from physical knowledge of the system, to infer the values of the parameters in the three equations model. What they tried to do is to sufficiently decouple the selected behaviour from the rest of the variability to justify the fact that simple dynamical systems may capture the dynamical properties of this mode, and to learn about the mode from palaeoclimate observations. Here, using the paleorecords, the calibration was applied inadequately: (1) fitting the parameters by maximizing a correlation coefficient (2) using the solutions of the optimization process as a set representing possible solutions of the predictions (and used as probabilistic estimates) (3) selecting the model with a very weak statistical criteria and unjustified threshold (0.7 for the correlation coefficient) : this is not a probabilistic forecast. A more adequate calibration method for the model as well as a more adequate verification and validation method of the predictive ability of the model are a must: any other choice must rely on a probabilistic treatment of the parameter and allow estimating uncertainty of the predictions. As stated in Cruficix (2012): "In a statistical inference process, the observations should be a plausible outcome or realization of the model. This makes sense only if the model has a stochastic component, which describes its uncertainties, limitations, and the noise that emerges from the chaotic motions of the atmosphere and oceans".

Two main approach : one can chose to handle the challenge of probabilistic forecasting long-term climate, or

1. Stochastic dynamical systems are used for inference on palaeoclimate time series.
2. Bayesian methodology, because it allows the integration of physical constraints in the form of prior distributions on model parameters. The Bayesian formalism is also naturally designed for model calibration, selection and probabilistic predictions (please, check Bayesian methods for selection and calibration of dynamical systems on noisy observations and the paper by Crucifix, 2012).

2.1 Critical comments and questions to be absolutely addressed

1. Neither the 100ky duration of ice ages, nor their saw-tooth shape were predicted by Milankovitch. Please check literature and update the knowledge.

Why the reviewer decided that we are not aware about these limitations of the classical Milankovith theory – we cannot even guess. Obviously our paper is not a review of the astronomical theory of glacial cycles. There are numerous publications, including those were AG was co-author, which present useful reviews of the current status of the understanding of glacial cycles such as Berger (2012), Past Interglacials Working Group of PAGES (2016), Berends et al. (2021). In our manuscript we devoted only one sentence to the Milankovich theory: "*The astronomical theory of glacial cycles (Milankovitch, 1941) postulates that growth and shrinkage of ice sheets is primarily controlled by changes in Earth's orbital parameters (eccentricity, obliquity and precession)...*". This statement is obviously correct. However, in others of our publications, we not only discussed these facts: "*One of the major challenges* *to the classical Milankovitch theory is the presence of 100 kyr cycles that dominate global ice volume and climate variability over the past million years "* (Ganopolski and Calov, 2011); *"Of particular interest is the transition between 1.25 and ~0.7 Ma ago, ..., from mostly symmetric cycles with a period of about 41 thousand years (ka) to strongly asymmetric 100-ka cycles"* (Willeit, et al. 2019), but also provided possible explanations for these facts.

As far as the "recommendation" to "*check literature and update the knowledge*" given by the reviewer#3 to the scientist (AG) who in 2011 received the EGU Milankovitch medal for "for his pioneering contributions ... to the understanding of the role of climate system feedbacks and the link between Milankovich forcing and global glaciation", published more than 30 papers directly related to mechanisms of glacial cycles and Milankovitch theory – such "recommendation" cannot be considered anything by rudeness. While the authors do not know who reviewer is, he/she knows the names of the authors even though is unable to spell them properly (the second author never published papers under the name "*Ganopolsky*").

2. This work is absolutely not a forecast work and nor a probabilistic forecast. This should absolutely be addressed and corrected. Without it, the paper cannot be published. This is a scenario based work, even not from a sensitivity nor a what-if scenario framework. as they only used three main scenarios (low, medium and high levels of starting point of carbon dioxide).

We used the term "forecast" in a very broad sense – under forecast we meant future simulations to distinguish from past climate simulations (hindcast). Obviously, no one expects it to be possible to make accurate climate forecast for million of years when it is not possible even for the next 100 years. But since of the term "forecast" can cause confusion, also in line with comments by reviewers #1 and #2, in the revised manuscript we will change "forecast" to "possible future scenarios".

3. This work embraced a method based on many unjustified simplifications and approaches. Please, Address the reasons and strong justifications why you accounted for the mentioned simplification (assumptions) of all the climate processes and the estimation of the parameters:
(a) The modeling approach: from line 264 "Finally, we approach the task of the selection of set of parameters P as a non-linear optimisation problem with equality and inequality constraints. We wish to find P to maximize the optimization target function(correlation criteria)" to line 289: This is not acceptable for forecasting, probabilistically
or not. how do you justify the selection of the best model, or calibration of parameters, while this is done via correlation: it is not probabilistic the way you did it. Neither it is an adequate one. It is like selecting the curve that suit you well given one aspect in the data, which might be linked to linear correlation! Why did not you considered any Least-Squares (Model Fitting) Algorithms? How about validation using scoring to select the best model,
there are many statistical criteria to select the best model, to fit and calibrate statistically and under constraints.
Honestly, either I did not understand at all what you did, or it is looking more like a patchwork using inadequate pieces! Especially seen in the following "See Appendix A for a discussion of the dependence of model performance on the choice of this time interval. To select parameters that will optimise correlation at the same time
as providing magnitudes in accordance to empirical estimations, an equality constraint is enforced: the maximum ice volume must be equal to 1 within a tolerance of 0.15 (in nondimensional units). Finally, the inequality constraint is given by Eq. (14).": this is really not acceptable.
(b) Validation set: Did you check the validity of the length of the time series used for calibration? how sensitive are the results given the the length of the time series used for calibration? how did yo find the optima length?
(c) Strong justification for not using appropriate probabilistic forecast models and adequate methods for calibrating the chosen one. No palaeoclimate record is dated with absolute confidence, so how do you account for the errors in the calibration data?

(d) Running multiple realizations by varying the model parameters: this is what is needed. but, this is not what you did! how did you considered that being probabilistic? What you did, is simply taking the solutions offered by the optimization process for multiple combinations of the parameters, choosing the sets that maximize the correlation with an unjustified threshold of 0.7! then using them as equivalent of multiple realizations of the predictive model to conclude about a probabilistic forecast! This is inadequate and inaccurate. what you did here, is just finding the
best set of parameters for your model. The way this has been done does not even give you the credible interval of the values for the parameters (and with an insufficient number of simulations as you run 1000, picked less then 400 and you have 9 parameters!).
Once you calibrate your model with the best set of parameters, verify it and calibrate it, then you should run an MCMC or any other sampling, to generate a set of probable realizations of your model, given the range of adequate values of the parameters, and a justified distribution for each parameters in the model.
Please check the literature for a proper way to do it including the optimal number of realizations which is far from 103.
(e) Results and from Figures: statements of results adequacy not validi. Figure 1: I really do not see that your predictions coincide with reconstructions. especially clear in figure 1-b!
ii. The magnitudes are not well reproduced at all.
iii. in appendix A: you have a correlation of 0.36... No comment!.
(f) The correlation level of 0.7, although arbitrary, guarantees a good fit to the paleo climatic ice volume record : this must not be used at a first place, it should certainly not be chosen arbitrary, and the figures do not show a good fit neither your correlation coefficients (using correlation at a first place is a problem in itself)
i. correlation is not an adequate criteria to assess goodness of fit in time series the way you did it
ii. it is not a good way to assess relation or association between time series (such as ice volume and CO2)
iii. correlation is insufficient by itself, and it assumes linear relations only.
therefore the comparison in between paleorecords and model output is weak, incorrect and incomplete.

Comment #3 contains numerous repetitions of the reviewer's believe that CR09 approach is the right one and ours is not. We believe, that the reviewer is fundamentally wrong.

The main reason why CR09 cannot be used for the design of the models suitable for "climate predictions" is the "no-analogue problem" or, in other words, the past is not the future. (See also discussion in the reply to Reviewer 1). The fact is that during the last 800 kyr for which reliable reconstructions of CO_2 concentration exist, CO_2 concentration was below 300 ppm, and most of time it was even below 250 ppm. At the same time, at present CO_2 concentration is already above 420 ppm and it is expected that at the end of the century it will be somewhere in between 500 and 1000 ppm. Assuming no negative net CO_2 emission in the future, CO_2 will stay for the next 100 kyr above 300 ppm even for optimistic 1000 PgC cumulative emission, which is higher than over the past 800 kyr. In the case of 5000 PgC emission, CO_2 will stay above 300 ppm for nearly 1 million years! Thus during the period of time in the future considered in our study, CO_2 will stay above the range its natural variability observed during the past 800 kyr. This is why it is not surprising that paleoclimate data are unable to constrain the most critical for future prediction parameter K (slope of critical CO_2-insolation relationship). After all, statistics is a not magic - it cannot extract from the data information which the data do not contain. Of course, we fully agree with the first reviewer that accurate paleoclimate reconstructions from a warmer climate state, for example late Pliocene and earlier Pleistocene would be very useful. Unfortunately, all CO_2 reconstructions prior to 800 kyr are very uncertainty and cannot be used to constrain model parameters. (To get an idea what "uncertain" means, one can make a look on Fig. 5 in Berends et al., 2021).

This is why we do not see any alternative to our approach, which, of course, is fundamentally different from CR09. The essential elements of our approach are:

1) We constructed a set of model equations based on general understanding of climate dynamics and the results of simulations with CLIMBER-2. This ensures that our simple model has stability properties and dynamical behaviour similar to CLIMBER-2. In particular, similar to CLIMBER-2, the simple model has two stable equilibrium states (glacial and interglacial), under orbital forcing simulates strongly asymmetric glacial cycles which are phase-locked to eccentricity and depend only weakly on the initial conditions, etc.
2) The anthropogenic CO_2 perturbation has been calculated using results of another EMIC (cGENIE)
3) We calibrated the model against paleoclimate data for the last 800 kyr and generated a large set of model realizations which simulate past glacial climates with the required accuracy
4) We rejected all model realisations which simulate glacial state at present
5) We applied a strict constraint on critically important parameters (slope of critical CO_2-insolation relationship) derived from CLIMBER-2 and thus arrived to a much narrower ensemble suitable both for past and future simulations.

Needless to say is that such approach represents a significant step forward compared to CR09 because CR09 described methodology for the calibration of the model suitable only for modelling of the past while we developed a model suitable for modelling past and future.

Furthermore, the reviewer heavily criticises the use of correlation in sentences (seen here and in other parts of the report) like: "correlation is not an adequate criteria to assess goodness of fit in time series the way you did it" "calibration using maximization of the correlation coefficient?! This really need to be explained and justified and proved working." "Correlation should not be used as a validation criteria!" "[correlation] it is not a good way to assess relation or association between time series" "correlation is insufficient by itself, and it assumes linear relations only" "Why did not you considered any Least-Squares (Model Fitting) Algorithms?"

The reviewer states that maximising correlation is not a proper fitting technique. This statement is incorrect. Please see Livadiotis and McComas (2013) who present the maximization of the correlation fitting method. Those authors show that the method is mathematically well defined under certain conditions and that it should be preferred over the classical least squares fitting in situations in which the data sets exhibit variations that need to be described, such as the variations that concern us here: glacial cycles.

4. Carbon dioxide curves: your choice of the evolution need to be justified. why should it be decreasing exponentially?

Why should we justify the use of the results of the well-established Earth system model published in a respected scientific journal? Moreover, these results are consistent with the previous findings. The reason for exponential decay of anthropogenic perturbation on very long-time scale is the removal of atmospheric CO_2 by weathering processes.

5. The relationship between critical insolation threshold for glacial inception and CO2 levels is known and must be analyzed using an appropriate sensitivity analysis.

This relationship is ONLY known from OUR paper (Ganopolski et al., 2016). This paper presents a single equation with a single set of numerical parameters. We do not understand how our formula can be "*analysed using an appropriate sensitivity analysis*".

6. Calibration/ Validation need to be done correctly (a) The validation part (crf. appendix A) is very weak. It has to be addressed with more adequate diagnostics for time series, especially graphical ones. (b) You must use a statistical criteria, more adequate to select the best model. large literature on that. (c) A sensitivity analysis or history matching plus an experimental design: would have been of high aid in this case where the hyperparameters have many constraints and we only know the range of the

parameters. designing a space filling set of combined parameters while constraining them in the space formed by all them. Run the optimization algorithm with only realistic combinations.

Regarding model validation, when producing a model with a predictive aim the gold standard is to evaluate the model predictive ability using previously unseen data (Stone, 1974). As we are dealing with time-series, the common procedure of randomly splitting the data into training and validation sets is inappropriate as it disrupts its temporal structure. We opted then to use an out-of-sample method, essentially holding out the last/earliest part of the time-series for testing (see for example Cerqueria et al., 2020). Given the cyclic structure of the paleo climatic data we are using, the only sensible option is to divide the information into 8 cycles (corresponding roughly to the 8 glacial cycles in the last 800 kyr). In the manuscript we reported (in Appendix A) the results for the model predictive ability when holding out 50% of the data for validation (i.e. holding out either the last 4 or the earliest 4 cycles). This is a quite stringent test for the model and a more adverse situation than the alternatives of holding out just 3, 2 or 1 of the cycles. We think the model validation methodology we employed is, therefore, adequate and not "weak" as the reviewer claims.

7. Please use the term "pacemaker" instead of "control" when referring to the astronomical forcing. The theory of ice ages has already evolved and, it is established that the astronomical forcing, especially for the assessing the particularity of the 100ky precession enigma (See Ditlevson and Crucifix (2017) On the importance of centennial variability for ice ages):"changes in eccentricity modulate the amplitude of precession peaks at a period of about 100 ka, but the spectrum of insolation time series do not contain an amplitude peak at this period. Source here (...) With this possibility in mind, the astronomical forcing is often prudently presented as the "pacemaker" of an internal oscillation rather than a primary "driver".". you can refer to the work by De Saedeleer, Crucifix and Wieczorek, https://dial.uclouvain.be/pr/boreal/object/boreal:119083 for a more systematic verification of the concept of forcing during ice ages.

In the "Crucial comment #7" the reviewer demonstrates the "knowledge" of the theory of glacial cycles by telling us to use the term "*pacemaker*" instead of "*control*" or "*driver*" when referring to the astronomical forcing and went further explaining that "*The theory of ice ages has already evolved*". While we fully agree that the theory did evolve, it evolved in the opposite direction to what the reviewer thinks. The term "pacemaker" in application to glacial cycles first appeared already in a paper by Hays et al. (1976) entitled "Variations in the Earth's Orbit: Pacemaker of the Ice Ages". Since then results of numerous simulations with physically-based models clearly demonstrated that orbital forcing is not just a pacemaker (this can mean essentially everything) but the real driver of glacial cycles. This was formulated in one of our paper as: "*Here... we demonstrate that both strong 100 kyr periodicity in the ice volume variations and the timing of glacial terminations during past 800 kyr can be successfully simulated as direct, strongly nonlinear responses of the climate-cryosphere system to orbital forcing alone...*" (Ganopolski and Calov, 2011). This result has been confirmed by numerous works done by Andre Berger, Ayako Abe-Ochi, Axel Timmermann and others.

8. The glacial inception problem

(a) In line 300, "(...) we analyse the critical insolation – CO2 relationship during glacial inception episodes for the different model realizations derived from Valid and compare them with Ganopolski et al. (2016)."Validating a work using results from another simulation, does not seem accurate to me.

(b) The glacial inception problem has been treated probabilistically and by using conceptual models. This study must be taken into account: refer to the work by Crucifix and Rougier 2009 on On the use of simple dynamical systems for climate predictions: A Bayesian prediction of the next glacial inception.

i. How do you position yourself comparing to the work by Crucifix and Rougier 2009?

ii. Why not to use the same idea for the modeling part?

As we explained already, we do not "validate" one model by another – we used the results of a complex model to construct and constrain a much simpler one. When measuring complexity in the length of program codes, CLIMBER -2 is thousand times more complex than our semi-empirical simple model.

CR09 presents a modelling approach which is not suitable for future climate prediction. We presented a model which is suitable for future projections. CR09 is based on fundamentally wrong model, in which glacial cycles represent self-sustained oscillations. Our model simulates glacial cycles the same way as complex models do. How we are supposed to use "*the same ideas*"?

9. « This approach, obviously, is not applicable for a possible future Antarctic and Greenland melting under high CO2 concentrations. This is why we do not consider future sea level rise above the preindustrial level and it is required that v_0 at any time» : Why? How do you justify that?

Not clear what is necessary to justify? Why our model is not applicable to Antarctica? Because the Antarctic ice sheet cannot be described by the same model as the Northern Hemisphere ice sheets, which is obvious. Or does the reviewer ask about potential problems related to neglecting future sea level rise? As we discussed in the response to reviewer #1, for the considered set of emission scenarios, this is not a serious problem (please see response to reviewer #1 for more details).

2.2 [Title] Need to be changed

The title must reflect the main goal of the paper. The paper is more on assessing the impact of anthropogenic CO2 emissions on the next 103 ky for recommendations on the the evaluation of geological disposal systems. The response to future environmental changes driven by a combination of natural (astronomical variations) and anthropogenic (fossil fuel emissions) forcing. Moreover, the only climate variable considered in this study is temperature (not representative of climate as a whole). Suggestion : impact of anthropogenic CO2 emissions on temperature in the next million years: assessment with a reduced-complexity model for glacial cycles.

We think the current title of the manuscript reflects its main goal. However, we understand that the use of the word "climate" might be too broad. In a revised version, we could consider modifying the title to address this comment.

2.3 [Section 1: Introduction] Need rewriting. It is not attractive nor well developed: the introduction must reflect the main subject.

Mainly: Rearrangement of the ideas from the main purpose of the paper then the necessary supporting facts! In addition, the introduction must highlight the advantage/choice of using this specific conceptual model in a
more relevant way. I think, here we need more details and justifications on the formulation of the framework/method/approach then reiterating about the ice ages and the Milankovitch theory (which can anyways be re/moved).

- Lack of consistency in the flow of ideas, lack of referencing on the main subject. It needs rewriting.
- I join Referee 2, to refer to the technical report by Lord et al.
- Please, refer to the work by Crucifix and Rougier (2009) and in a more profound way the cited paper Cucifix (2012). Of course, you must add complementary papers in the same line as these two.
- It is well established that climate change is a human activity induced.

May be drop lines from 25-45 in the introduction, and use them as supporting facts for supporting the following points in order: by explaining

(1) the goal which is more related to "the challenge of the permanent storage of the radioactive waste" and " The evaluation of geological disposal systems in response to future environmental changes, driven by a combination of natural (orbital variations) and anthropogenic (fossil fuel emissions) forcings (e.g. Lord et al., 2016) is, therefore, mandatory" so start the introduction with line 47 (while adapting the text of course).

(2) Why we need to consider a model for glacial cycles and why we must include, in the simulation study, natural and human induced factors : human activity induced impacts on climate change has a long term impact.

Use lines [25-45] as supporting facts Or use it to support the justification of the calibration part in the methodology section, Line 85 when discussing the ice ages.

(3) Then proceed with line 54 starting from "However, these timescales are(...)".

(4) Please, add a more adequate review of literature related, specifically, to the subject of analyzing or assessing the impact of carbon dioxide concentration variations (under scenarios) on the stability/evaluation of geological disposal systems.

(5) when you say "to this end" : I do not see how you account for the "quantitative probabilistic assessments." in your proposed model. why not to announce already your approach here in a concise way. because, conting for the quantitative probabilistic assessments is not part of the defined/designed predictive model, the " reduced-complexity process-based model of the coupled climate – ice sheets – Carbon cycle evolution, whose only external forcings are insolation and cumulative anthropogenic CO2 emissions."

Please specify that the interest is the evolution of temperature. Justify why (linking it to the main subject of the paper which is "the challenge of the permanent storage of the radioactive waste and The evaluation of geological disposal systems").

(6) Please, position more adequately your contribution. It is not clear from the text. Please, point out the lack in the literature (If there is so) and the breakthrough of your study and advantages of using your approach/choice of model and parameterization. For instance, what was the outcome and the lack(s) in the work of Archer and Ganopolski (2005) based on Paillard's conceptual model? And, why did you chose here the simulator based Earth model from Lord et al.?

(7) Please, refer to the most up to date theory of the astronomical forcing instead of Milankovich. check the paper by Curifix and Rougier (2009) for a detailed explanation and the theory and its history.

(8) Note that Milankovitch's theory is missing the dynamical aspect of climate's response and that the Glaciologist Johannes Weertman (J. Weertman, Nature 261, 17 (1976)) is the one who addressed the evolution of ice sheet size and volume by means of an ordinary differential equation (ODE), "thereby opening the door to the use of dynamical system theory for understanding Quaternary oscillations" (Crucifix and Rougier, 2009). This need to be highlighted in your paper and used as a reference as your work is about modeling using ODEs.

We agree that a re-organization of the Introduction section to highlight first the main goal of the manuscript is a good idea. We also agree that we need to position our contribution more adequately among the existing literature.

We do not agree with the statement "specify that the interest is the evolution of temperature", as it is not. We will highlight that our model was designed for simulations of future glacial cycles while global temperature is a very useful diagnostic which can become necessary in other potential applications of our model.

Please, see comments done earlier on the Milankovitch's theory.

2.4 [Section 2: Model and datasets] Form

Start with the set of all equations where equation of temperature will be first, then then explain the need to parametrize each of them.

So:

• Start with Subsection 2.1 Please, shall you design a flowchart to show all the parts of the modeling framework. addd a table with all the parameters to be inferred during the calibration process add a table gathering all notation, acronyms and definitions of variables, put here or in the appendix. Introduce the set of equation first (3 equation while starting with the temperature one). define the parameters, use lines 204:208.

• Follow up with subsection 2.2: details of the the equations then explain and explicit each equation, its meaning, goal, parametrization....and here you need just two subsubsections. (one for ice and the other for CO2, no need for temperature as it is in sec.2.1)
• Follow up with subsection 2.3: explicit the constraints... and so on (subsection 2.5 in the draft paper) start from line 215.

• Subsection 2.6 is used for describing the data used for validation: please, move it to section 3 (model performance).

We see no advantage in modifying the order in which the equations of our model are presented. We agree on modifying Table 1 to better highlight which parameters are to be inferred and to add another table gathering all the notation.

We believe the equations, with the explanation of each term, are sufficiently well presented in the current form of the manuscript (except for the temperature equation, which could be improved).

We agree that subsection 2.6 could be moved into section 3.

Comments on the method: critical to be addressed

1. How do you account for uncertainties in the observational data while calibrating the model?
   We agree that in the current version of the manuscript there is no deep discussion about potential problems with the plaeoclimate reconstructions used. We will address this shortcoming in a revised version.
   To account for the uncertainties in the paleodata is that we consider as valid all the solutions with correlation between paleo and modelled ice volume higher than 0.7.

2. How do you justify the choice of 0.7 as acceptable for the correlation coefficient?
   The selection correlation higher than 0.7 is designed to filter only those solutions which reproduce the ice volume behaviour in the last 800 kyr reasonably well (considering also that the paleorecord used for this variable is of course not perfect).

3. How do you justify the formulation of changes in temperature as a linear combination of global ice volume and logarithm of CO2 concentration? This part need a more thorough justification, explanation, development.
   The first term in the temperature equation represents the direct link between ice volume and global temperature anomalies (more ice volume in the NH is associated with lower global temperatures). The second term is explained by the fact that the radiative forcing of CO_2 is proportional to the logarithm of CO_2 concentration. We will clarify this better in a revised version.

4. calibration using maximization of the correlation coefficient?! this really need to be explained and justified and proved working.
   Please see comments done before regarding the suitability of the use of correlation in this context. In a revised version we will explicitly explain this choice.

5. How do you qualify your calibration/modeling/prediction method?
   The meaning of this questions is unclear. The evaluation of the methodology from the point of view of its potential predictive skill was evaluated in Appendix A. Based on those results we qualify the methodology as having a satisfactory ability also when used in predictive mode.

2.5 [Model performance]

This part has to be done appropriately, once the modeling part is fixed and an appropriate calibration method is selected. This part should be applied statistically to verify and validate the calibrated model. Comparison of the model predictions with paleoclimate data (reconstructions) should be assessed within the calibration process. The length of the calibration time series should be assessed (assess the

predictability of the model given the length of the time series). Correlation should not be used as a validation criteria!
Please check literature for validating models calibrated for time series: this is what you need to learn and know, to work in this subject and write your paper.

Model performance from a predictive point of view was evaluated in Appendix A. We already discussed that the use of correlation as a performance metric is justified in this case, as there are major variations that need to be reproduced in order for the model to be of utility.

2.6 [Conclusion]
The conclusion has to be adapted and rewritten with all the paper.
Just a note on: "It is also clear, however, that even though there is a high level of agreement in the solutions' trajectories during the past 800 kyr, their paths tend to diverge for the future indicating that the past does not perfectly constraint the future evolution of the climate – ice sheets – Carbon cycle system." : I do not think this is absolutely necessary to mention: we know that and this experiment is not needed, the statement either.

We think this is one main aspect from our results and deserves to be mentioned in the conclusions.

3 Secondary comments
To help correcting/adapting/improving the work/paper, it would be beneficial to the authors to check definitions/methods/literature (in a general framework and then for time series, and in paleoclimate field) on the following:
• conceptual models
• predicting vs forecasting
• probabilistic forecast
• (probabilistic) sensitivity analysis
• simulating using scenarios
• decision making based on scenario assessment
• probabilistic calibration of models based on time series
• verification and validation of calibrated models (set of diagnostics)

[General] Please,
• Use one verb tense for adequacy. Also, either direct form with the use of "we" or indirect with the one other verb tense. Such as in lines 70 to 73.
Agreed, will be modified.
• Remove the expression "can be found": where ever it is in the text, it has to be changed into an active voice verb, such us is +adequate verb (displayed, shown, ...).
Agreed, will be modified.
• Refer to the technical report of Lord et al., on the same topic "Modelling changes in climate over the next 1 million years"
Agreed, will be included.
• Refer to the work by Crucifix and Rougier 2009 on the probabilistic modeling of climate change on the glacial inception.
Agreed, will be included.
• The only variable that is important to address the problem of storage is temperature. I suggest to keep any other figure (ice and CO_2) in the supplementary material.
We do not agree with this statement. In fact, the prediction of the timing and magnitude of the next glaciations is the main output from our model useful for decisions related to the disposal of nuclear waste in deep geologically-stable rock formations. Global temperature is a useful diagnostic which can become necessary in other potential applications of our model.

• Remove "please" in line 162 and 194, and if any other in the text.
Agreed.

• change "orbital forcing" into "astronomical forcing" wherever it occurs and adapt the text accordingly. For instance, a sentence such as "The orbital forcing f(t) depends only on astronomical parameters (eccentricity, precession and obliquity) " in line 206 is unnecessary.

We will remain using the term "Orbital forcing" as it is standard in the specialised literature.

[line 28] "Antarctic ice core records also show" : to be consistent with line 25 and the statement "Numerous paleoclimate records show (...)", avoiding the use of "also" would be preferable.

Proposition:

During this period, atmospheric Carbon dioxide (CO2) concentration fluctuated nearly synchronously with the global ice volume, and CO2 concentration during glacial times was up to 100 ppm lower than during preindustrial 30 time, as shown in Antarctic ice core records (Petit et al., 1999, Lüthi et al., 2008).

Agreed.

[line 32] "Earth's orbital parameters". These are astronomical parameters.

"Orbital parameters" is the term usually employed in the specialised literature, we will continue to use it.

[line 35] May be more adequate using "supported" instead of "confirmed" as per verifying a theory by the aid of a reduced order model and/or a simulator which is not enough to infer knowledge for conforming a theory but verifying it or validating an aspect of it with a set of verifications (Reductionism based knowledge inference especially based climate simulators cannot be used as a tool to confirm anything).

Agreed.

[Lines 70-73] Need rewriting, adapting the verb tenses. Please stick to one verb tense for adequacy. Also, stick to one form passive or active ("we"). Better : if you use a direct simple style with present tense.

Agreed.

[210] Correct "in a good (see discussion below) agreement" to "in a good agreement (see discussion below)"

Agreed.

[225] use "condition" or "constraint" instead of "criteria" in "The last imposed criteria" for consistency with the text.

Agreed.

[236] why do you use "limitation". in all this section you are introducing constraints. use the term "constraints" everywhere and count them as being 7 in total.

Agreed.

References

Abe-Ouchi, A., Saito, F., Kawamura, K., Raymo, M. E., Okuno, J., Takahashi, K., and Blatter, H.: Insolation-driven 100,000-year glacial cycles and hysteresis of ice-sheet volume, Nature, 500, 190–194, doi.org/10.1038/nature12374, 2013.

Archer, D., and Ganopolski, A.: A movable trigger: Fossil fuel CO2 and the onset of the next glaciation, Geochemistry Geophysics Geosystems, 6, 7, 10.1029/2004gc000891, 2005

Berends, C. J., Köhler, P., Lourens, L. J., & van de Wal, R. S. W. (2021). On the cause of the mid-Pleistocene transition. Reviews of Geophysics, 59, e2020RG000727.

Berger, A., and Loutre, M. F.: Modeling the 100-kyr glacial-interglacial cycles, Global and Planetary Change, 72, 275-281, 10.1016/j.gloplacha.2010.01.003, 2010.

Berger, A.: A Brief History of the Astronomical Theories of Paleoclimates, 107-129, in Climate Change, A. Berger et al. (eds.), Springer-Verlag, Wien, 2012.

Cerqueira, V., Torgo, L., & Mozetič, I.: Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, *109*(11), 1997-2028, https://doi.org/10.1007/s10994-020-05910-7, 2020.

Ganopolski, A. and Calov, R.: The role of orbital forcing, carbon dioxide and regolith in 100 kyr glacial cycles, Clim. Past., 7, 1415–1425, https://doi.org/10.5194/cp-7-1415-2011, 2011.

Ganopolski, A., Calov, R., and Claussen, M.: Simulation of the last glacial cycle with a coupled climate ice-sheet model of intermediate complexity, Clim. Past, 6, 229–244, https://doi.org/10.5194/cp-6-229-2010, 2010.

Ganopolski, A., Winkelmann, R., and Schellnhuber, H. J.: Critical insolation-CO2 relation for diagnosing past and future glacial inception, Nature, 529, 200–204, https://doi.org/10.1038/nature16494, 2016.

Ganopolski, A., and Brovkin, V.: Simulation of climate, ice sheets and CO2 evolution during the last four glacial cycles with an Earth system model of intermediate complexity, Clim. Past., 13, 1695-1716, 10.5194/cp-13-1695-2017, 2017.

Gregory, J. M., Browne, O. J. H., Payne, A. J., Ridley, J. K., and Rutt, I. C.: Modelling large-scale ice-sheet-climate interactions following glacial inception, Clim. Past., 8, 1565-1580, 10.5194/cp-8-1565-2012, 2012.

Hays, J. D., Imbrie, J., & Shackleton, N. J.: Variations in the Earth's orbit: pacemaker of the ice ages. Washington, DC: American Association for the Advancement of Science, 1976.

Livadiotis, G., & McComas, D. J.: Fitting method based on correlation maximization: Applications in space physics. *Journal of Geophysical Research: Space Physics*, *118*(6), 2863-2875, https://doi.org/10.1002/jgra.50304, 2013.

Paillard, D.: The timing of Pleistocene glaciations from a simple multiple-state climate model, Nature, 391, 378–381, 1998.

Past Interglacials Working Group of PAGES (2016), Interglacials of the last 800,000 years, Rev. Geophys., 54, 162–219, doi:10.1002/2015RG000482.

Stone, M.: Cross-Validatory Choice and Assessment of Statistical Predictions, Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 111–133, https://doi.org/10.1111/j.2517-6161.1974.tb00994.x, 1974.

Willeit, M., Ganopolski, A., Calov, R., and Brovkin, V.: Mid-Pleistocene transition in glacial cycles explained by declining CO2 and regolith removal, Sci. Adv., 5, 8, 10.1126/sciadv.aav7337, 2019.

Wolf, E., et al.: Interglacials of the last 800,000 years, Reviewes of Geophysics. 54, 162–219, 2016