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 and Maasch (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.

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').

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.

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.

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 behavior 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.

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).

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 10^3.

(e) Results and from Figures: statements of results adequacy not valid

i. 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 CO_2)

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.

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

5. The relationship between critical insolation threshold for glacial inception and CO_2 levels is known and must be analyzed 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.

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.

8. The glacial inception problem

   (a) In line 300, *“(...) we analyse the critical insolation – CO_2 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?

9. « This approach, obviously, is not applicable for a possible future Antarctic and Greenland melting under high CO_2 concentrations. This is why we do not consider future sea level rise above the preindustrial level and it is required that $v \geq 0$ at any time» : Why? How do you justify that?

Crucifix, M., Rougier, J. On the use of simple dynamical systems for climate predictions. Eur. Phys. J. Spec. Top. 174, 11–31 (2009). (DOI).

_Ditlevsen, P., & Crucifix, M. (2017). On the importance of centennial variability for ice ages. Past Global Changes Magazine, 25, 152-153.

_Crucifix, M. (2012). Oscillators and relaxation phenomena in Pleistocene climate theory. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370(1962), 1140-1165.

_De Saedeleer, B., Crucifix, M., & Wieczorek, S. (2013). Is the astronomical forcing a reliable and unique pacemaker for climate? A conceptual model study. Climate Dynamics, 40(1-2), 273-294.

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 CO_2 emissions on the next 10^3 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 CO_2 emissions on temperature in the next million years: assessment with a reduced-complexity model for glacial cycles.

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* CO_2 *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.

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 CO_2, 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).

Comments on the method: critical to be addressed

1. How do you account for uncertainties in the observational data while calibrating the model?
2. How do you justify the choice of 0.7 as acceptable for the correlation coefficient?
3. How do you justify the formulation of changes in temperature as a linear combination of global ice volume and logarithm of CO_2 concentration?

   This part need a more thorough justification, explanation, development.
4. calibration using maximization of the correlation coefficient?! this really need to be explained and justified and proved working.
5. How do you qualify your calibration/modeling/prediction method?

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.

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.

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.
- 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, ...).
- Refer to the technical report of Lord et al., on the same topic " Modelling changes in climate over the next 1 million years"
- Refer to the work by Crucifix and Rougier 2009 on the probabilistic modeling of climate change on the glacial inception.
- 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.
- Remove "please" in line 162 and 194, and if any other in the text.
- 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.

[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 (CO_2) concentration fluctuated nearly synchronously with the global ice volume, and CO_2 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).

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

[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).

[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.

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

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

[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.