

Review of the article “Using Deep Learning to integrate paleoclimate and global geochemistry over Phanerozoic time.”

General comments:

Spatially resolved modeling has gained popularity in recent years and has proven to be a very effective tool for assessing paleoclimates. Due to the sparse data (modeled or from proxies) used in assessing fluxes through time, these models are limited. Building on SCION's already important development in this domain, this contribution develops a method that makes the model even more reliable. At least two reasons make this study very valuable and interesting:

1. It provides a framework to improve the accuracy of surface processes used in biogeochemical models and demonstrates that 10Myrs is sufficient which gives valuable information to the community for future studies.
2. This method reduces the global computing costs of running models over geological timescales, which are currently a major limiting factor in many research projects.

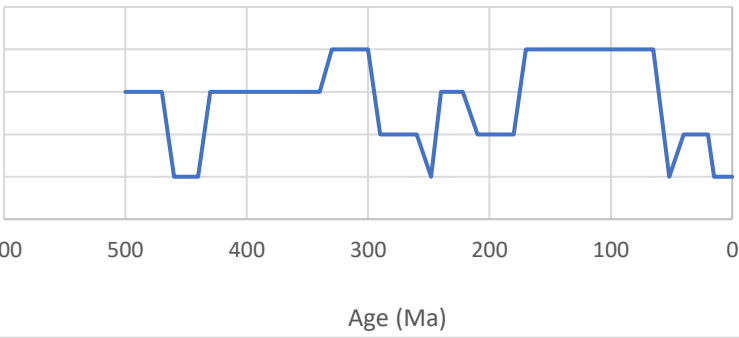
It was a pleasure to read this manuscript, which is of high quality. The scientific significance is excellent as the sparse data available is a limiting agent in the paleoclimate modeling domain. The scientific quality is very good as studies are carried out to test the reliability of this method and so on different timescales and targets (paleogeography vs runoff) and discussed in some detail. The scientific reproducibility is good as the method is explained thoroughly and models are available to download. The presentation quality is good as the figures are very relevant to the text and illustrate it well.

However, the PaleoDEM validation revealed some serious issues with interpolating over 40 Myrs, yet the authors used it on GEOCLIM/SCION with intervals up to 55 Myrs. Additionally, Figure 4 shows artificial landmasses created with a 10Myr timestep that may pose a problem for climate modeling, mostly for models including oceanic circulation. Therefore, some caution on the use of this method should be highlighted a bit more thoroughly in the text.

Specific comments: *(the numbering applies to the preprint version of the manuscript available online)*

<u>Lines</u>	<u>Comment</u>
title	Phanerozoic time sounds a bit odd to me, why not use the normal word for it: “eon”?
26-27	“Species distribution”, you don’t mention this point in the main text, maybe it will be better placed in the conclusion as an opening?
38	Godd�ris et al. (2023) is quoted but can’t be found in the reference list.
44	It is mentioned 22 continental configurations. However, in both GEOCLIM and SCION, only 21 frames are cited (not assigned to any reference though). What and when is this 22 nd ?
47-49	From what I understood the ITCZ is more or less forced by the model, is that what you mean? However, the ITCZ's shape and location will be determined by the paleogeographical configuration and be modeled by the GCM, along with the areas of extreme weathering. In most cases, it is not the other way around. Would it be possible to reformulate this sentence?
60 - 63	When I read this manuscript, what for me is the heart of this study is the method and validation parts, the latter is not mentioned at all in the introduction.

68	4.5°*7.5° might be too coarse to display important paleogeographic features such as island arcs that might have a great impact on climate over some periods of the Earth's history (see Ribeiro et al., 2022 or Marcilly et al. 2022)
68-69	"Roughly evenly spaced" You then mention later in the text (l 137) that some spacing is 55 Myrs. Moreover, it is never mentioned in the text what is the model used for these continental reconstructions. If it is indeed the one used in the original GEOCLIM (Goddéris et al., 2014) then is it the maps from Blakey (2007)? (As in Nardin et al. 2011) I think this study will gain from referencing the paleogeographic models in a better way because they are at the base layer for climate modeling and therefore are extremely important.
71	"Original FOAM", which version of the model do we speak about?
75	"Wide spacing in time between ...datasets": the time spacing is mentioned but still no numbers are given → maybe it will be nice to give some numbers for the reader to know which scale of spacing we are talking about?
95-96	You mention shifts in FOAM are due to the reorganization of landmasses, yet no real plate tectonic/paleogeography studies are quoted here. It will be nice to have the study presenting the reconstructions behind FOAM quoted.
142	Maybe "the" should be "a" PaleoDEM dataset. PaleoDEM is a quite widely used term and not restricted to the work of Scotese.
154	The problem with downscaling is that it often results in an overestimation of the exposed land ratio which will mean more area available for weathering for the biogeochemical model. I know it's difficult to run GCM with a finer grid. <i>(It was just to raise awareness as this is not strictly related to the subject of your study. I understand the point here is to demonstrate the reliability of the method which I think is very well done in this study)</i>
227-228	I'm not sure what you are trying to say here: that the synthetic and real maps (Scotese and Wright, 2018) have a better fit together compared to Scotese & Wright (2018) and Marcilly et al. (2021)? If so, it might need some reformulation. What period of comparison are we talking about here? Can you give an estimate of this discrepancy, in % error for example?
274 - 275	Having issues with small landmasses is quite serious because they often display high runoff (Goddéris et al., 2014) and therefore host high weathering.
294-296+ § 4	This is interesting because other models such as GEOCARBSULF (Marcilly et al., 2021) and GEOCLIM (Goddéris & Donnadieu, 2017; Goddéris et al, 2014) have this spike which is attributed to a change in climate sensitivity in GEOCARBSULF for example. I'm confused about how the frames are now interpolated; in this section, are we back to the first part where you interpolate following the spacing of the maps which are roughly evenly spaced"? how much time in between two frames? It is a bit confusing after the validation part with the 3 different spacings. So, you demonstrate that the accuracy with spacing greater than 10Myrs is reduced using this method and yet you use it with intervals up to 55 Myrs? Is that not a problem? I don't think you should draw any conclusions with intervals over 10 Myrs which if it is indeed the FOAM runs as the one in Goddéris et al. (2014), actually covers the majority of the Phanerozoic.

	<p style="text-align: center;">Time spacing between 2 maps from Godderis et al. (2014)</p>  <p>This is where it becomes complicated for me to understand because the extreme warmth of the Permian Triassic extinction is probably shorter than 10 Myrs so can you actually see this signal? in Cao et al (2022) the interval considered is 253-247 Ma, roughly 254-250 Ma in Yang et al. (2019) for example.</p> <p>How can you see short signals such as the P/T warming but not the Ordovician cooling (Hirnantian) for example? (Which has been attributed by many to be caused by changes in paleogeography (e.g., Nardin et al., 2011).</p> <p>Why quoting Wu et al. (2023) the article in the reference list is not about the PT boundary.?</p>	
301	The fate of South China will also depend on the chosen reconstruction and downscaling process as for the lower Triassic in Fig 9 South China seems well emerged but in the reconstructions of Marcilly et al. (2021) the land area available for weathering is very small.	
305 - 307	You should also mention that the timestep between two “base” reconstructions is greater than 10 Myrs and the accuracy is therefore reduced.	
323	<p>“20 Myrs apart” are they though? From Godd�ris et al. (2014) they seem more 30 to 40 Myrs apart for the majority.</p> <p>“Can be applied “vs “should aim to run climate models at least every 10 Myrs” (I333) Therefore, the recommendation made for further studies is not respected in this very study → a bit of mixed message. Maybe this sentence should be rewritten?</p>	
334	The conclusion is very well structured and easy to read. Maybe it will also be worth mentioning here that the method fails to reconstruct short-lived events (greater than 1Myrs though) such as the Hirnantian glacial event. Even though they have large climatic consequences.	

Comment related to figures.

<u>Figure</u>	<u>Comment</u>
2	Concerning the graphs presenting SSIM and 2D correlation: Whatever the frame interval considered it seems that there are two periods of increasingly low performance. It’s difficult to read for sure the ages but I would say between 430-420 Ma and 250-210 Ma. What can cause this?

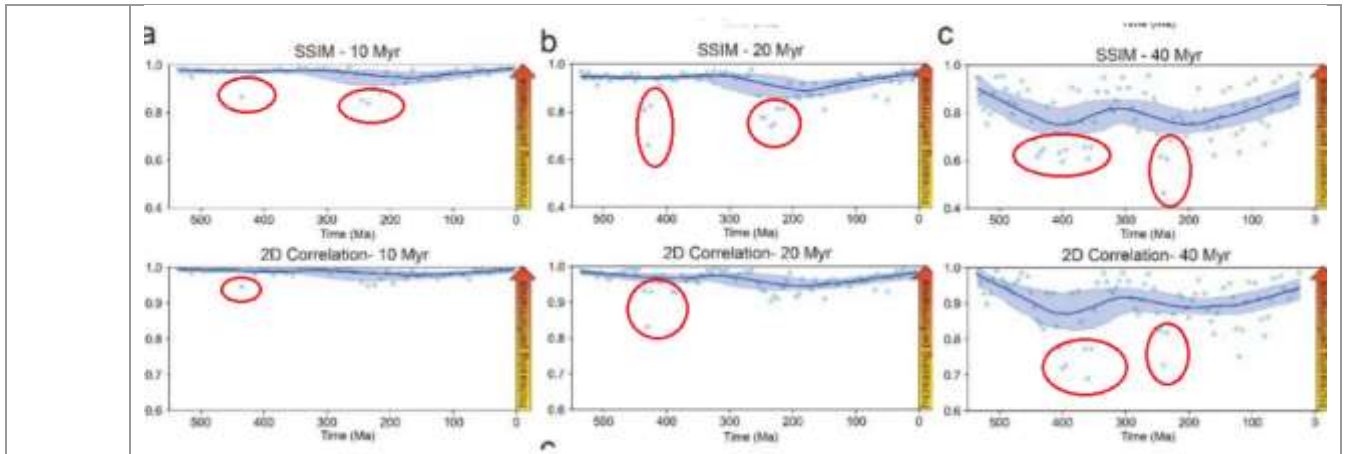
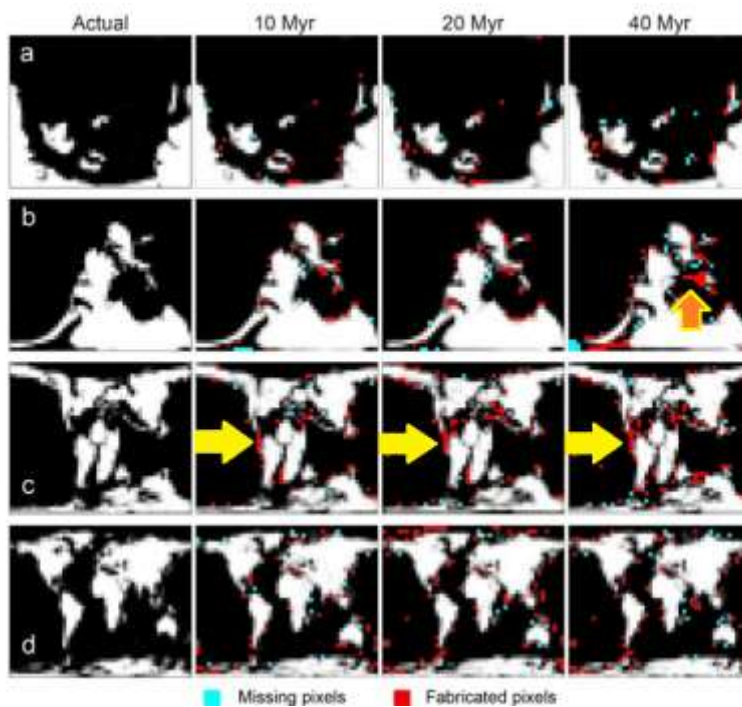


Table 1 Title: Missing the a in “evaluation”

The synthetic maps at 105 Ma (yellow arrows) worry me a bit because it's ok (not ideal though) for running GCM simulations with FOAM because it doesn't have a proper oceanic circulation module but with other GCMs such artifact will represent an issue. Can you comment on that?

You don't mention it in the text but this creation of land over South China (orange arrow) will create a big issue for the assessment of weathering fluxes at it is well known small, isolated landmasses are hosting a lot of runoff and therefore weathering (mostly at the equator). It will be nice to highlight this point even if you already mentioned that the 40 Myrs step is less accurate. It will actually illustrate this point.

4



5

In both the runoff and temperature graph, deeper time runs seem to have a better correlation between CO₂-Temperature and CO₂-runoff: How do you explain that?

6-7

Those two figures are quite crowded, is it possible to select the most representative graphs and put the other ones in the appendix?

<p>8</p>	<p>It will be nice to see an estimation of the “accuracy” of the method on this figure. Maybe highlight the periods where base maps are closer to each other and so lead to more accuracy. This way the reader can directly see which periods are more reliable than the others.</p> <p>In the text you mention, that with this method, the timestep is reduced to 1Myr so we should see the signal of more short-lived events. However, here the Hirnantian cooling is totally hidden and instead there is even an increase in CO₂ and temperature. Can you comment on this?</p>
<p>9</p>	<p>The figure highlights the increase in runoff and weathering in central Pangea during the Late Permian /Early Triassic in mid-latitude Pangea, but this is debated and evidences such as large extend of evaporites deposits suggest quite arid conditions (Scotese maps below (DOI:10.13140/2.1.2757.8567.) or Cui & Cao (2021) https://doi.org/10.1002/gj.4123).</p> <p>Arid conditions are unlikely to result in intense weathering. Can you comment on that as well?</p> 