Response to Reviewer 3

This paper presents a procedure to create surrogate trajectories of climate model ensembles. The authors provide tests on a set of CMIP6 simulations and discuss the sensitivity to two key parameters of the procedure.

5   I have no reason to doubt that the authors know what they do. My main concern with the paper is that I neither understand the general picture nor the details.

>>We are sorry that our paper turns out to be so opaque to someone not familiar with the topic. We thank the >>reviewer for the open mindedness and fairness with which the paper was evaluated. We find the comments >>very useful and we have
10   made an effort to to make our work better understandable by a larger audience.

My first concern is on the format of the paper and its suitability for ESD. The abstract, introduction, and conclusions are written by and for IPCC insiders, as the authors use a lot of IPCC jargon, which is obscure to most human beings, including me. This style of writing seems to go against the interdisciplinary nature of ESD. Not only the paper
15   does not report new understanding of the climate system, but the authors do not discuss that their procedure might help do so (or how). Another example is the use of the term "emulator" or "emulation". Of course, this remark is not limited to this manuscript. I yet have to see a reasonably clear definition of what is called a "climate emulator". For some authors, an emulator is a regression between some predictand
20   variable and a predictor. Here, it is obviously something else, that looks akin to analog modelling. Making a proper bibliographic search could help relate the procedure described in the manuscript to existing work, which might not appear in the IPCC reports. The notion of "creating new scenarios" is not clear. The IPCC seems to use SSP scenarios, which are relevant for the economy. What the authors do is
25   obviously something else. So, using this terminology might be confusing. The simple (acknowledged) fact that the emulation procedure cannot produce relevant GHG (or any forcing fluxes) should plead against the use of "creating scenarios". My understanding is that the procedure creates surrogate trajectories that are constrained by GSAT values. Why should those trajectories be called "scenarios" in
30   the IPCC sense?

1

>>These questions are extremely useful in pointing at the need for clarification and shedding jargon. We have >>attempted to do so, and in the process also clarified the emulator purpose, and its product. Throughout we >>have attempted to carefully phrase emulation as the emulation of ESM output, not emulation of scenarios, >>which is a contraption of the actual meaning that thanks to the reviewer comments we are now aware of >>being potentially confusing. You will find the revised manuscript extensively edited for clarity and when not >>completing doing away with it, defining the IPCC type nomenclature.

My second concern is that the procedure description seems inappropriately vague. Ideally, I should be able to reproduce the procedure by reading the manuscript (provided I have access to the data). The first step (l. 148) suggests that *one* time series of GSAT is created for each model by dumping together all ensembles, scenarios, etc. ([…] "the time series is made […]"). I guess/hope that the authors do differently. The fourth step (l. 157) is not clear: what is a target scenario? The authors allude to "target scenarios" in several places, but do not define what those are. I believe that the authors could design a diagram that explains how the procedure works. In practice, I understand that one needs to know the target scenario (i.e., have GSAT data). Hence, I do not understand how the authors can reconstruct "unknown" scenarios (e.g., SSP2) from just SSP1 and SSP5, which suggests that intermediate scenarios can be deduced from two extreme SSP scenarios. This might be true, but I would like to understand this miracle (at least for me).

>>We have evidently failed to communicate the basic set up of our problem. We have extensively rewritten the >>methods section, and we have now added a diagram that we hope should clarify the steps in the >>construction and the outcomes of the algorithm. Thank you for the suggestion of including such graphic.

My third major concern is on the results or the performance tests. The authors seem to be happy with the results reported in Figs 1-8. Indeed, the "emulated" time series are close to "targets" (whatever how the targets were designed). But is this desirable? The GSAT time series have no decadal or interdecadal variability (which might due to the procedure itself). This is not discussed, but I would doubt any procedure that

2

creates trajectories that do not yield long term variability are so useful, or really account for climate variability (e.g., the so-called butterfly effect). For me, the SOI results are "good" by construction, since they are excerpts of existing simulations. How would this emulation procedure be able to emulate changes in ENSO variability, which would be a key issue for impact modelling? My feeling is that the simulated trajectories give overconfidence about (the lack of) climate internal variability. The conclusion that this procedure can replace numerical model simulations hence seems overconfident.

>>Again, we think there is a fundamental problem of communication at play, as we are proposing a way to >>produce ESM output according to a new scenario stitching together existing ESM output, so STITCHES >>retains all the internal variability characteristics of the original ESM output. Of course, we are aware of and >>we discuss limitations in this regard, due to the idea of stitching together windows of existing simulations of >>9 year length, and due to the fundamental assumption that most variables are scenario independent in their >>behavior (this addresses the concern about ENSO variability changes) so their characteristics, as they would >>be produced by the ESM if running the scenario that instead is being emulated, are preserved as long as the >>algorithm matches the corresponding global warming levels. In particular, for ENSO variability changing at >>higher warming levels the idea is that we would sample such behavior for our emulation by sampling ESM >>output at high waring levels. Again, this assumes that the change in variability is essentially scenario >>independent and all that matters is the warming level.

Minor issues

In the search of nearest neighbors in the (T, dT) space (step 5, l. 165), are there different weights on T or dT, in the distance definition?

>>At the moment they are used with equal weight in a Euclidean distance, but the algorithm could be >>tuned/modified to define a distance that weighs one more than

the other. It's indeed an important >>possibility, and we have mentioned that in the text.

In step 6 (l. 170), what is a "pointer"?

>>We have rewritten extensively the description of the algorithm so we hope that now the terms will make >>better sense. We call pointer the identified archived experiment/time window that will give us content for a >>specific segment of our emulation.

l. 211: "(see 1)", what is "1"?

>>Apologies, the word "Table" was accidentally forgotten.

Figure 1: I can't read the labels on a printed version of the manuscript.

>>Also Reviewer 1 alerted us to the need of increasing the fonts. We will.

I feel that there should be a separate section that describes the experimental set ups, tests, etc.

>>We are hoping that by having clarified the rationale and functioning of the algorithm the present structure >>works better.

Figures 2-4: the captions should only keep descriptive statements, not comments that already appear in the text.

>>We have cleaned these up.

Equation (1) (l. 360): all symbols should be introduced. What is the bar? I think that $\hat y$ should be the synthetic and y the truth, not the other way around, as suggested in l. 362. E2 is certainly not a ratio of variances (but a ratio of standard deviations). The denominator of E2 should be: $<(y - \bar y)^2>$ (no $\hat{}$).

>>Apologies, the reviewer is absolutely right that the equation and its terms are not rigorously presented and >>explained. We had also made a mistake in the formula,

where the denominator of the second component >>should be the same as that of the first, as the reviewer pointed out. We have fixed the typos, defined the >>quantities and symbols and corrected the text. We have also eliminated the bar notation, which had the >>same meaning as the angle brackets.

5 In conclusion, my feeling is that the manuscript would be much more appropriate in GMD, which incidentally has a better impact factor than ESD. Of course, this decision is left to the authors and the editor.

>>Thank you for helping us identify these shortcomings. We will attempt to make our paper clearer and >>hopefully more interesting to the audience of ESD, which

10 we would still prefer to pursue. In particular, we >>think the impact research community would be interested in our proposal, and we believe ESD could reach >>that community more easily than GMD.