Earth System
Dynamics

Discussions

EGU

Interactive comment on "Systematic Correlation Matrix Evaluation (SCoMaE) – A bottom-up, natural science-based approach to identify Indicators" *by* Nadine Mengis et al.

Nadine Mengis et al.

nadine.mengis@concordia.ca

General Comments:

This paper describes a method to identify indicators of environmental change that have statistical measurability, political and societal relevance, and scientific consistency. Indicator identification follows a novel 3-step process, based on the construction of a correlation matrix and significance of the r values. They use an Earth System Model of intermediate complexity run with 3 scenarios (historical, RCP 4.5 and 8.5) and 16 sensitivity perturbations in an example of their methodology.

While this paper was interesting and the concept was of indicators being used to describe natural and forced states is sound, there were some major issues that I feel need to be addressed in the introduction and methods sections.

Thank you for your generally positive evaluation of the applied method. We hope to have addressed all of your suggestions mentioned below to your satisfaction. We feel, that your comments helped us to improve the paper.

Specific Comments:

First, what is the common way to identify/construct indicators? Why would this methodology be new, novel, different, better, etc? Their argument is one-sided and more background into other methods is needed. It is implied that expert judgment is the only criteria for indicator selection.

We added two references on indicator selection to give more back ground on this topic and point out why we think this should be improved: "Working group II of the Intergovernmental Panel on Climate Change (IPCC) (Houghton et al., 2001) used [global mean surface air temperature (SAT)] as the main climate change indicator, due to its predominance in the existing literature and its large scientific consistency as such." (page 2, lines 12-15) "However, selecting one indicator, while disregarding the other is a normative choice Krellenberg et al. (2010), which can (unknowingly) be biased by e.g. technical knowledge (Rametsteiner et al., 2011). In this study we want to introduce a bottom-up indicator selection method, that uses statistical information about variables in addition to expert judgement, attempting to reduced this bias in the selection process." (page 2, lines 29-33)

This type of methodology was not compared across models, hence results are strongly model specific. Also, the methodology could only be applied if the ability to examine model sensitivity exists through perturbation runs. This would limit applying this method to specific models or would be far too computationally expensive to pursue.

Based on your comment and also the comments from the second reviewer, we decided to add a section in which we explain in more detail the set up of the example. Here we explain that comparing our perturbed parameter simulations, is similar to comparing a multi-model ensemble. This would be a computationally doable application. However, also the application of this method to temporal correlations is possible. (See new Section 2 on page 3)

The rationale for using 11-year averages for calculating delta was not given. I can assume that because these are global averages and the model has low internal variability (referenced from page 14 line 20) that 11 years is sufficient? 10-30 year periods are more common. Also, for the calculation of the historical period deltas, what years were used? The title on Figure 3 would suggest that some linear method was used to calculate delta? I could not find specific years used.

The fact that the model has low internal variability was the reason for the 11-year averages. We identified the years in Figure 2 and 3 more clearly. Thank you for pointing this out.

Additionally, 5% significance level is extremely low. What is the rationale for using such a low significance value? Why not use an absolute threshold for r instead? Or inflate the ensemble used for the correlation calculation by using more than 1 realization from each perturbation? Indicators from the historical period are interesting in that there are likely very few variables that have a clear signal or significant trend. Is this why the very low 5

Using a 95% confidence that a correlation is significant, is a standard value often used in scientific studies, and was in our opinion appropriate for the assessment of all three scenarios, where varying signal strengths are present. Given that we only have 16 simulations, this is a quite strict criteria. However, we wanted to make sure that there is a real signal in the correlation between the variables reaction to our parameter perturbations.

Cluster analysis: Why would this method be preferred over more standard methods of PCA/EOF analysis or step-wise regression to group variables? Can the results be compared? The emphasis seems to be on non-redundancy of variables, but what is the value of non-redundant groupings?

The point we wanted to make here is that we might be able to obtain very similar information from the variables within one cluster, e.g. sea ice and temperature, these are then redundant and there is more value in regarding variables from another cluster, which we would possibly have neglected due to our possibly unknown bias (see page 2, lines 29-31). The value of non-redundancy is also within not un-proportionally stressing one result while reading multiple correlated indicators. We agree with the reviewer that our method should be compared to other ones. However, we feel that such a comparison is too much for this publication and we are currently working on a follow-up study in which we compare our results to a PCA, to see how similar the results obtained by those methods are.

In the discussion section (Page 12, line 27-28), it is stated that "our results do suggest that a comprehensive assessment of future climatic states needs a re-evaluation of the ad-hoc chosen indicators, due to changes in prevailing climatic responses." This is an interesting statement in that it implies that the historical indicators are often used to determine future states. Are there previous studies to compare this to? If there were more details in the introduction, this statement would make more sense and have a stronger impact.

The study by Rametsteiner et al., (2011) points a little bit into the direction, that a reassessment of sustainability indicators with time is important, in case new knowledge is obtained about the Earth system, but in their case also in case of changes in societal preferences. We point to their study in the introduction now (see page 2, line 30), but we think, that the finding of changed correlations in future high CO_2 and temperature scenarios and the according implications for

indicator selection, is one of the novel points in our study.

Minor comments: Page 2, line 23: The citation uses "respectively" but it is not needed. **Removed.**

Page 3, line 2. Is there a reference for this model? There are many references in the appendix section, but there should be one here as well.

We added a reference.

The correlation matrix is constructed by taking correlation coefficient of 2 variable deltas under 16 perturbation experiments. I am assuming you are using Pearson correlation coefficient (r). This should be stated at least once in the methods.

We added a cleared explanation of the calculation of the correlations.

The caption of Figure 1 could use more detail, and the figure itself could be generalized. If this paper is about the method (and not the model) it should be much cleaner and clearer. For example, if the 16 perturbation runs are not discussed in the methods section, why are they shown in the figure? This critique also applies to Figure 2.

We hope to have addressed these issues by adding a new section 2, which explains the experimental set up as well as the single parameter perturbation, before these figures appear.

The naming structure of the data was confusing. Prefixes A, O, F, and L were used. I am assuming that A was either atmosphere or absolute, F was flux, and O was ocean? There were also suffixes of O or L, which are ocean and land, and N or S (North and South?). While I understand that these variables were used to showcase the method and as such the naming conventions were not important, it was hard to interpret the results based on variable names.

We added an explanation of the prefixes at least in Figure 2 and 3. Hoping that this improves the basic understanding, otherwise we made sure to have the ex-

planation of the variables names, when ever they are mentioned in the text. Also we added table S1 to the appendix now.

In the abstract, there are 3 non-optional steps, but in the manuscript Step 3 is listed as "optional".

In fact, we already regard step 3 as optional, so we changed the text accordingly.

Page 5 line 25, should there be forward slashes in the variable names?

We changed the text to make it more clear that we we referring to two variables here.

Figure 5, the title of the x-y correlation 'quilt' has the word "periods". Should this be changed to scenarios instead? It is not clear what is meant here. This isn't a plot of the significance of the changes in periods, but more of the significance of the r values between variable deltas?

We changed the heading to read scenarios now, thank you for pointing this out. Also we edited the figure caption hoping this makes it more clear now.

Page 15, line 16: what is meant by "natural science-based assessments"? "Natural" is used throughout the manuscript, but I am not sure it is needed.

This was referring to a natural science versus social science assessment. We changed the phasing now to science-led assessment, following the phrasing used in Ramentsteiner et al., 2011.

The discussion in the Supplementary section is very interesting, but I would argue that it doesn't belong in this paper. However, it could and should be used in a follow-up paper.

Thank you, we removed everything that is not directly referred to in the main article and may publish some of it later.

References: Houghton, J., Ding, Y., Griggs, D., Noguer, M., van der Linden, P., Dai, X., Maskell, K., and Johnson, C.: IPCC Third Assessment Report: Climate Change, 2001 (TAR), Working group II: Impacts, Adaptation and Vulnerability, Chapter 19, 2001. Krellenberg, K., Kopfmüller, J., and Barton, J. R.: How sustainable is Santiago de Chile? Current performance, future trends, potential measures. Synthesis report of the risk habitat megacity research initiative (2007-2011), Tech. rep., UFZ-Bericht, Helmholtz-Zentrum für Umweltforschung, 2010. Rametsteiner, E., Pülzl, H., Alkan-Olsson, J., and Frederiksen, P.: Sustainability indicator developmentâĂŤScience or political negotiation?, Ecological indicators, 11, 61–70, 2011.

Please also note the supplement to this comment: https://www.earth-syst-dynam-discuss.net/esd-2017-72/esd-2017-72-AC1-supplement.pdf

———————————————————

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-72, 2017.