Manuscript reviewed: "An open-access CMIP5 pattern library for temperature and precipitation: Description and methodology"

Authors: Cary Lynch, Corinne Hartin, Ben Bond-Lamberty, Ben Kravitz

Journal: Earth System Science Data

Date: 3 March 2017

The authors have constructed a dataset that includes scaling patterns for a wide number of global climate models (GCMs). These patterns can be utilized in creating future projections for such forcing scenarios that have not been explicitly simulated by the models.

Recommendation: The dataset is accessible but the manuscript needs substantial revisions.

1 Data evaluation

- Uniqueness: rating = 2
- Usefulness: rating = 1
- Completeness: rating =1
- Data quality: rating = 2
- Comments: The methods used for creating the data (linear regression etc.) are not new but the effort has been substantial and has required much resources. Scaling patterns are available for 41 GCMs at monthly, seasonal and annual level. For controlling purpose, I downloaded one file; the file was easily accessible. This example file was readable by all three softwares/commands that I proved: grads, ncdump and cdo.

  One problem is that the files must be downloaded one by one. Regarding the large number of files, a wget tool similar to that in use in the CMIP5 data bank would be useful for those users that are interested in a large ensemble of models. Such a tool might be developed later, and this issue is not any reason to delay the publication of the database and article.

  It is impossible for me to check the correctness of the data included in the files; we have to trust that the team has done their work properly in this respect.

2 Manuscript: Main comments

The authors have used much space to show that the regression method outperforms the delta-change method in producing scaling patterns. However, these details do not constitute necessary information for the users of the database. For example, exploring the dependence of delta change patters on the reference and scenario periods is quite far from the main focus of the paper. (If this discussion is shortened, some more detailed comments presented below may become irrelevant.)

There is one very good additional reason to provide scaling patters just for the regression method and not for the delta-change method: the delta-change patterns are much more straightforward to calculate by the users themselves than the patterns based on the regression method. This might be mentioned in the paper as a motivation for selecting this method.

I find that sections 1–2 are moderately good but the authors should have edited section 3 much more carefully.

1. Page 1, line 6-7: "This paper presents patterns from all CMIP5 models for temperature and precipitation...". In fact, the paper only evaluates the temperature patterns but no results are given for precipitation. Even so,

precipitation files are available in the repository as well. Accordingly, a brief evaluation of the precipitation patterns should indeed be included in the paper (pattern minus original model output).

2. In the repository, patterns are given for 41 GCMs, but in the manuscript, for some reason, the evaluation of the methodology and patterns is limited to 12 GCMs only. No motivation for this choice is given. Studying a wider ensemble of GCMs would give a more robust picture on the differences between the methods, scenarios, etc. In particular, it is strange that no GCM from China (e.g., BCC-CSM1-1) has been included in the ensemble!

3. As a criterion for the similarity of the multi-model mean patterns, you use the Student's t test (page 4, line 30 onwards). As the sample is small (12 GCMs), the test only reveals very large differences between the means. Accordingly, the t value remaining below the significance threshold does not provide any strong evidence for the actual similarity of the means. Text (on page 5, lines 23–24 and 29; page 6, lines 11–12,...) would also need revision.

   Perhaps, it is not so essential to study in detail whether the differences in the patterns are statistically significant. It is trivial that the pattern depends, to some extent, on the RCP scenario, reference/scenario period, pattern-scaling method, etc. Whether these differences are statistically significant or not, largely depends on the size of ensemble. The larger the number of GCMs, the smaller differences would be significantly different. You have selected to provide patterns founded on the linear regression, and, in my opinion, any inter-method differences do not invalidate that choice.

4. I do not find the comparison presented in the first paragraph of section 3.2 relevant. The outcome is fairly trivial: if one compares dissimilar periods, the GMT change is different as well. The similarity of the changes under the RCP8.5 scenario appears fortuitous. In particular, the values presented in Fig. 7 do not give the rate of GMT change, since the length of period in L21C/L19C and the 21st century is far from the same.

5. Caption of Fig. 6 is unclear. I presume that "modelled change" refers to the multi-model mean change derived directly from the original model output while "predicted pattern" stands for the pattern calculated by one of the pattern-scaling methods. This should be stated more clearly. Moreover, if one wants to elucidate the performance of the scaling methods, it would be more logical to show the difference (scaled value) minus (reference, i.e., the original model output). In the present representation, positive values evidently refer to an underestimation (page 6, line 14), which makes the interpretation difficult for the reader. Finally, I did not understand "when GMT change = 1 (without any unit!)"; perhaps the idea is that the maps have been normalized to correspond to an 1°C increase in GMT. In that case, do you have any explanation why the global mean of the difference is not equal to zero?

6. In Table 2, there is something that I do not understand. In Fig. 6, the differences are far < 1 everywhere. How can you obtain RMS differences that are much larger than the differences occurring at any individual grid point? Regarding these large RMS differences, how can you state that "Nevertheless, Table 2 indicates that the both methodologies do well emulating actual model output."

3 Specific comments

1. Page 1, lines 10 and 14: The magnitude of the actual temperature increase depends on the radiative forcing. It would be relevant to give here normalized responses in °C/°C rather than the absolute responses.

2. In the www address given in the abstract (http://doi.org/10.5281/zenodo.235905), I only found the annual mean patterns. By contrast, in the address stated in section 6 (https://github.com/JGCRI/) there was a full manifold of files.

3. In the abstract, it should be explicitly stated that the patterns in the repository are based on the least square regression method rather than on the delta-change method.

4. Page 4, line 27: In order to reproduce the time series of TL_{MS}, ε should be three- rather than two-dimensional (a function of time in addition to the latitude and longitude).

5. On page 6, there are sentences that are hard to understand, e.g.: "To evaluate performance of each pattern methodology, accuracy was based on how well the patterns approximated the linear GMT change of 1°C simulated by each GCM."

6. Page 6, line 18–19: "Overall, it appears that the regression pattern scaling method underestimates the relationship between global temperature and local temperature, but the degree to which it overestimates the relationship is small (< 0.08°C)." Some interpretation should be given for this. Is this difference in global mean temperature change induced by the use of a simple climate model?

7. Page 7, line 12–13: The nonlinear evolution of temperature is related to the retreat of sea ice.

8. Page 7, line 16: R^2 evidently refers to the square of correlation coefficient; this should be mentioned explicitly. Is the correlation calculated with respect to time?

9. Page 7, line 20: "despite lower local and global trends". This might be removed. The ratio may well be large regardless of the magnitude of the numerator and denominator.

10. Page 7, line 21: Over oceans in winter, the rate of warming is large before the ice melt. After the melt, the large thermal inertia of the sea water tends to decelerate warming.

11. Page 9, line 3: I studied one file by the ncdump command, and in that file the period for historical climatology was 1961–1990.

12. Page 8, lines 13–14: "particularly when strong mitigation is employed later in the simulation". The idea is not clear for me.

13. In the Figures, colour hues are very close to one another, and thus it is hard to infer the actual values of the field at some specified location. For the largest values, it would be good to use some other colours than red and blue. Alternatively, you might add some contours to facilitate interpretation.

14. In Figs. 4 and 6, some other colour than white should be used to emphasize areas with statistically significant differences.

15. Fig. 3: Is the variance calculated with respect to time? Standard deviation might be more illustrative than the variance. Moreover, the variability of temperature is very different in the different parts of the world. Therefore, it would be relevant to give differences in std in a relative sense (e.g., $(\sigma_1 - \sigma_2)/\sigma_2$). This would reveal where the differences actually are pronounced. After seeing the differences in rms in relative terms, it is possible to evaluate the statement "Differences in variance across epochs were also small" (page 5, line 19) and whether "these relatively small differences in variance (or std?) between epochs were not likely to affect the resulting temperature patterns".

4 Minor comments

1. Page 2, line 13: "adaptability of additional predictors" is difficult to understand.

2. Page 2, line 20: Does "... vary between scenarios and for other climate variables...." mean "... vary between scenarios and climate variables"?
3. Page 2, line 31–33: "In linear regression, only the error term is assumed to have a normal distribution, so it is likely that climate extremes would yield high error terms." Please explain this more explicitly.
4. Page 4, line 12: "a 30-year reference epoch of 1986–2005". Only 20 years!
5. Page 4, line 15–17: The sentence should be rephrased.
6. Page 4, line 18: 1861–1990 should be 1861–1890?
7. Page 5, line 2: "ensemble variance". Variance within or between the ensembles? Please specify.
8. Page 5, line 22: "in the observed period". You are dealing with model simulations rather than observations.
9. Page 7, line 6: Should the unit be °C/°C rather than °C?
10. Page 8, line 6: I did not understand "The local to global fit is strong".
11. Page 8, line 8: Give some examples of potential additional predictors.
12. Page 8, line 16: "epoch trends"?
13. Page 8, line 26: Please define GitHub.
14. Page 9, line 6: 165 kB to 1 MB.
15. In the pdf file of the manuscript, list of references is given twice.
16. In Table 2, are the rms differences global?
17. In the caption of Fig. 4, specify the RCP scenario.
18. Fig. 7: The unit is different for change (°C) and trend (°C/year).
19. In Fig. 9, consider showing the standard deviation rather than variance; give the unit of the quantity in the caption.
20. Fig. 10, caption: Replace "R-squared" by "the square of the correlation coefficient".
21. Fig. 11: Is this "Ensemble mean annual average"?