Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-170-RC2, 2016

Geoscientific Model Development Discussions

Interactive comment on "Exploring global surface temperature pattern scaling methodologies and assumptions from a CMIP5 model ensemble" *by* Cary Lynch et al.

Anonymous Referee #2

Received and published: 29 August 2016

This article focuses on two simplistic pattern scaling approaches (delta and linear regression method) which are used to create additional future forcing scenarios for the use in impact assessment, adaptation and mitigation. They successfully demonstrated that the chosen time period (epoch) for the delta approach does not significantly change the resulting pattern. Moreover, they show that pattern scaled fields have the largest errors at high latitudes. Even though the regression method was shown to better match the modeled trend, it has its limitations when it comes to consistency across scenarios.

I recommend to only consider this submission after the following major revisions have been implemented. Additionally, the presentation and labelling of the Figures should

be improved to make it easier for the readers to understand what is shown.

General Comments Most of the results shown in this study are 30-year averages. But at the same time the authors highlight the importance of this study for impact assessments. Are the authors able to demonstrate skill of the pattern scaling methods at impact relevant temporal scales?

The model data was regridded to a T85 resolution (approx. 1.4° spatial resolution), even though Table 1 reveals that several models have resolutions which are significantly coarser than the resolution which is regridded to. Regridding to a finer resolution is strongly discouraged, as models clearly have no information at this scale due to parameterizations. This can lead to large errors in the model fields which is particularly harmful for the purpose of impact assessments. I recommend to regrid to the model with the coarsest resolution (in this case, MIROC-ESM which has a spatial resolution of 3°). Consider using the CDO function called remapcon (Jones, 1999) for this task.

Having multiple models per institute is not the only reason why the assumption of model independence is not valid. An important point is shared code or even whole model components. See, e.g., Abramowitz, G., & Bishop, C. H. (2015). Climate model dependence and the ensemble dependence transformation of CMIP projections. Journal of Climate, 28(6), 2332-2348. The way the authors deal with the issue of model interdependence is not appropriate. Simply selecting the "best" model per modeling center does not get rid of the internal dependence structure. It is for example known that the ACCESS models are based on the HadGEM2 atmosphere and both of those models are part of the selected CMIP5 subset (see Table 1). See, e.g., Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. Geophysical Research Letters, 40(6), 1194-1199. Model selection based on performance only is not recommended as one should ensure a high degree of independence at the same time. Pairwise error correlation can for example be used as a metric for model dependence, as used in Bishop, C. H., & Abramowitz, G. (2013). Climate model dependence and the replicate Earth paradigm. Climate dynamics, 41(3-4),

885-900.

The authors decided to use NCEP1 for the validation of their model ensemble based on annual and seasonal climatology. For proper model validation, I highly encourage the use of more than one observational/reanalysis product as observation products can vary significantly, even for surface air temperature. Especially in this case, I would not recommend the use of NCEP1 (a 1st generation NCAR product) as more recent reanalysis products are available. Here are some examples: JRA55, MERRA2, CFSR (3rd generation NCAR product), ERA-Interim. Why did the authors choose a reanalysis product over a gridded observation product? There are several gridded global observations available, especially for monthly temporal resolution. For example: HadCRUT4, Berkeley BEST, GISTEMP, NOAAGlobalTemp. Reanalyses should only be used for evaluation if no gridded observation products are available, because they contain model output. Therefore, the model ensemble might not be truly independent of the chosen reanalysis product and a comparison is not fair.

The conclusion section does not contain any comments on the limitations/potential caveats of this study and on future work. Please add a few lines.

Specific Comments Abstract The abstract assumes that the reader already knows what the linear regression and delta methods are. Please add a 1-2 sentence explanation of each method as it is otherwise hard for the reader to understand what was done. The last sentence of the abstract says that the patterns will be used to examine feedbacks. Where in the main text are feedbacks discussed?

Introduction Page1, Line 22: Typo: “It is used to...” rather than “It can is used to...”. Page 2, Line 3: No need to introduce the abbreviation IAM as it is not used later on in the article. Page 2, Lines 8-10: The word “increase” is used 3 times in this sentence. Try to change that (e.g., use warming rather than increase). Page 2, Lines 10-11: Please clarify this sentence. The pattern obtained with the pattern scaling methods is just a mean pattern and contains no information about the variability. This sentence

therefore does not make any sense to me. Page 2, Line 26: Please elaborate on why the reference period is expected to matter for assessing future change. Isn’t it rather the length of the reference period that is important?

Data/Methods Page 4, Line 1: Which and how many ensemble members per climate model were used for the analysis? Table 1: Add information about the chosen ensemble members. Replace “HadGEM-ES” with “HadGEM2-ES”. Page 4, Lines 8-10: Elaborate why the ensemble mean is not expected to be skewed when using model output that is not independent. Model replication automatically skews the mean of a distribution towards the most replicated one. Page 4, Line 11: Many impact studies need precipitation information as well as temperature information. Could this analysis be expanded to include total precipitation rather than a single variable? Page 4, Lines 14-15: The authors stated above that the assumption of model independence within the CMIP5 archive is not valid. In this case, how do the authors justify the use of the model spread as a measure of uncertainty? Having dependent models automatically leads to a bigger model agreement, which should not be regarded as an indication of robustness. Equation 1: It would be helpful for the reader to add the dimensionality of each element of the equation. So, DP and ΔTL are 2D fields whereas ΔTG is a number. The same could be done for Equation 2. The authors calculate the delta pattern for each model and then use the ensemble mean /median for their analysis. It would be interesting to compare those results from the pattern obtained when using the ΔTL and ΔTG of the mean/median of the ensemble (instead of using the individual models and then calculating the mean/median). Page 5, Line 9: What is a gridded time series? Is that simply 3-dimensional data (time x lat x lon)? Equation 3: Why is the “ˆ” symbol on top of β being used? Should this also be done in the text which describes this equation? Does “n” refer to the number of time steps? Please add that to the description of the equation. Page 5, Line 21: What does LSR stand for? Page 5, Line 26: Add “Herger et al, 2015” as a reference, as they also conducted a PCA and found that the global warming explains the largest percentage of variance across scenarios (see their Figure 1).

Results/Discussion Figure 1/ Page 6, Lines 5-9: Why does Figure 1 show absolute temperature even though anomalies are used for the pattern scaling approaches? Differences between the model ensemble and reanalysis are expected to be large if one considers absolute values rather than anomalies. I recommend adding multiple observational products to this figure for validation (and replace NCEP1 with a more recent product). Moreover, I recommend terminating the time series of the observational product in 2005, as there is no meaningful comparison between observation and model data later on. Moreover, the authors could add the ensemble mean (dashed) in addition to the median (solid) to make it clearer that the shading is symmetrical about the mean. Figures 2-4, 6-11: To make it easier for the reader to quickly understand what is shown in the Figures, the authors should add labels (ANN, DJF, JJA, rcp, time period,...) outside the panels rather than just having it in the caption. Moreover, add the unit to the colorbar and rotate the colorbar labels in Figures 3, 4, 6-11. Figure 2: The unit should be °C/year. Moreover, as the plots are symmetrical about the 1-1 axis, the authors do not necessarily need to show the lower half. They could potentially show the results of rcp8.5 in the lower left corner and the results of rcp4.5 in the upper right corner to reduce the number of subfigures from 6 to 3. Table 2: After reading Lines 17-21 (Page 6) it is not clear to me what additional information Table 2 gives compared to Figures 2 and 3, which already show the change in mean and trend. If Table 2 is kept, please add the abbreviations M21C, L19C to the caption. Moreover, the unit of the trend should be °C/year rather than just °C. Figure 4: Add (a) to the top and (b) to the bottom plot to make a clearer distinction between those 2 plots. Then refer to them properly in the text (4a in Page 6, Line 23 and 4b in Line 26). The caption says “Ensemble mean global air temperature delta method pattern difference based on differences in historical (top) and future (bottom) reference periods”. However, the reference period in both plots are within the historical period (1861-1890 or 1971-2000). Why are they referred to as historical and future? Figure 5: Put the months and their x-ticks in the center between the 2 boxes to make it clearer that 4 boxes belong to 1 month. Moreover, the caption does not explain what the difference

between the boxes on the left and right are. Which one is the global air temperature difference and which one is the trend? Page 7, Line 1: Adjust: “The different methods used to quantify global change are first exam-ined by comparing future minus present change (left box) to the 21st Century projected linear trend (right box, Figure 5).” Page 7, Line 10: Add a space between “(Figure 6)” and “and”. Figure 6: Add a space between “(bottom)” and “in” in the caption. Figure 7: Maybe mention in the caption that the scale in the colorbar is not linear (also Figure 9). Moreover, Figure 7 could potentially be incorporated into Figure 6 with the help of stippling/hatching for significant differences. Having both the differences and significance in one Figure would make the interpretation of the results easier. Page 7, Line 23: “regression pattern scaling method” rather than “regression pattern method”. Page 7, Line 29: A high comma is missing after “drier”. Page 8, Lines 18-19: Parentheses are for clarifications and referencing and should be avoided in this case. See, the following article by Alan Robock in EOS: http://onlinelibrary.wiley.com/doi/10.1029/2010EO450004/abstract Acknowledgments Page 9, Line 12: The ESGF link does not seem to work (privacy issues). Please remove the “9” from the link.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-170, 2016.