Response to reviewers
gmd-2016-170
Lynch et al.

We thank all referees for their thoughtful comments, which have substantially strengthened this manuscript. Per the _GMD_ instructions, we respond below to all comments and questions.

Original reviewer comments in italics. Responses in blue. Specific revisions in maroon.

Reviewer #3:

1. Given that the point of PS is to explore the spread of projections and to do so at local/regional scales, there is far too much on the ensemble mean and too much on the global mean temperature changes. This is mostly irrelevant to PS. The global mean temperature is an input to PS, rather than an output, so it is very odd to spend so much time analysing it and few conclusions seem to be drawn from it.

Response:
We use the ensemble mean GMT change in the discussion of epoch differences only, and when we construct patterns, the averaging is done after patterns are generated for each model. A thorough examination of GMT allows us to conclusively say that epoch choice does not significantly alter the resulting pattern despite differences in GMT trend and absolute value.
However, we agree that such discussions are supplementary and are not directly related to any conclusions about pattern scaling. As such, we have moved former Figure 1 and 5 to the supplementary section, as well as taken out Table 2 and Supplementary Figures 3 & 4 to make our conclusion about epoch choice more concise.

---

2. The work lacks a clear and logical overall framework. It needs to consider what factors will influence the patterns diagnosed, how will each diagnosis method be affected by these factors, how can the pattern differences be explained by these factors? Factors to consider are (i) internal variability versus the signal of forced climate change, (ii) nonlinear dependence on GMT and (iii) scenario dependence.

Response:
We agree, and these are areas which we have added further analysis, or discussed in terms of caveats and/or future work. We have made significant changes to address the lack of clarity and logical framework of this study. Below we address the three factors.
For (i), we included a figure of the r^2 of the between local and global temperature to show that GMT accounts for a significant portion of the variance found in the local temperature (>90%). We include a figure that shows the detrended variance for each future scenario and a signal to noise ratio plot. These plots show that for temperature patterns, the signal is very strong and the variance is relatively small, and internal variability does not significantly alter the resulting pattern.
For (ii), we did not do additional analysis on non-linearity. We have discussed methods to examine nonlinear relationships between local/ global temperature, in particular spherical harmonics and Fourier decomposition but analysis of this is beyond the scope of this study.

However, for temperature, the figures mentioned above indicate that at least for the analyses presented here, nonlinear dependence on GMT is small. We realize that if we were to examine precipitation or other climate variables, we will have to explore more complex statistical measures.

For (iii), Reviewers 1 and 2 had similar concerns. To further explore this issue, all figures included both scenarios, regardless of result section. We have found that the resulting scenario does not change the resulting pattern because the pattern is simply a relationship. In our analysis, significant differences in patterns between the two scenarios are primarily due to relatively large differences in the local temperature change as compared to the GMT change, but these are also areas where the local variability is high. An additional figure shows this ratio difference between scenarios.

Also for (iii), in a signal to noise analysis and we do not find significant differences between patterns despite differences in signals (either local or global). We have included a figure that shows the difference between the signal to noise ratio between the two emission scenarios to emphasis this point, and it is discussed in the 'Scenario Difference' section.

---

3. Then consider how method choices affect these: do you use initial-condition ensemble means for each model or just take a single run (this isn't stated), do you apply a time-filter (e.g. running mean or non-overlapping means) before applying the regression approach and how long is the running mean?

Response:

We used only one realization from each model, which was not stated in the text. We have added a statement to the "Climate Models" subsection to address this.

A time filter was not used. We chose to keep a 30 year climatology for the delta patterns, as this appeared to be the standard in the delta methodology.

The length of epochs used should not alter the resulting pattern. Barnes and Barnes (2015) found that for temperature, one-third the length of the time series is ideal, and for a 100 year time series the standard thirty year epoch length is sufficient.

---

4. *Then consider how these factors may explain the different patterns found. For example, in the final paragraph of the conclusions it states "the GMT temperature sensitivity is stronger when using a lower forcing scenario because... changes in GMT have a stronger effect on local temperature". This cannot be true across the board because the global mean of the local temperature changes has to equal the GMT change regardless of level of forcing, by definition. Therefore this explanation fails, and it should be reconsidered in terms of what regions exhibit stronger (apparent) local sensitivity between scenarios and why is this more apparent for regression versus delta methods?*

Response:

This statement was inaccurate and has been removed. As discussed in point iii above, the difference between scenarios is due to the difference in the local temperature change as compared to the GMT. We now discusse how the difference between scenarios in the regression method appears to be the result of a local sensitivity and variance differences, and have added a figure to show this difference, as well as a figure in the Supplementary section that shows 21st Century detrended variance.

---

5. There are significant flaws throughout the manuscript, possibility suggesting some fundamental misunderstanding of pattern scaling or at least they could lead to misunderstanding by the readers. Some of these flaws are listed below.
The units of Figs. 4, 6, 10 and 11 are incorrect (and units for SI Fig. 2 are not given), given as °C while it is evident from equation (1) for the delta pattern and equation (3) for the regression pattern that the patterns (DPMS or BETAMS) are dimensionless (or equivalently are °C/°C, expressing local temperature change per degree of GMT change). This error casts doubt on all the PS patterns shown and analysed here.

Response:
Thank you for pointing out these embarrassing mistakes, which we have corrected in the text and figures.

---

6. The basis of the manuscript is the assumptions that underlie the delta and regression methods used in the literature to diagnose the PS patterns (note that Osborn et al., 2015, is listed incorrectly here as using the delta method but it uses the regression method, explained in great detail).

Response:
The mention of Osborn et al, 2015, was indeed incorrect. The correct citation should be Osborn et al, 2009. It has been corrected in the text, and the Osborn et al, 2015, reference has been correctly attributed to the regression method.

---

7. Where the stated assumptions come from is not properly explained and the stated assumptions are in fact not all made by the pattern diagnosis methods. Under "Assumptions" it is stated that the delta method assumes that anthropogenic forcings do not modify internal climate variability. This may be an assumption of how PS is subsequently applied in some cases to produce a future projection (something that is not considered nor explained in the current manuscript), for example simply adding a PS change to an observed timeseries, but PS does not have to be applied this way (see Osborn et al., 2015, for a PS application using a GCM prediction of enhanced internal climate variability under anthropogenic forcing) and it is certainly not an assumption of the way in which the pattern is diagnosed from the GCM data in the first place.

Response:
The introduction has been extensively edited, and we have added the references for the underlying assumptions. We also believe that the assumption of stationarity in relation to variability is important in pattern creation as well as scaling patterns. While changes in variability may not change the GMT change and local change, they may influence errors in pattern fit. We have edited this section to more clearly state this.
The issue of variability changes across epochs was not properly explored in the manuscript. We have rectified this oversight by including a figure in the Supplementary material that shows spatial differences in variance across multiple epochs. This figure replaces the previous figure (S2).
We also do not use the patterns to generate a time series, as we are only looking at patterns, and will not be scaling the patterns in this manuscript.

Revised text:

In pattern scaling, the underlying assumption is that responses to external forcing and internal variability are independent, implying that anthropogenic forcings do not modify the internal variability of the climate system (*Mitchell*, 2003; *Lopez et al.*, 2013), but this premise is not always true (Screen, 2014). Changes in variability may introduce estimation errors in pattern fit, and in practice, estimation errors introduced through this assumption at the global scale are small but can be large enough at the regional scale to mislead adaptation decisions (*Lopez et al.*, 2013).
Another assumption in pattern scaling methodologies is that local change scales proportionally with GMT change, and that the relationship is stationary over time (*Mitchell*, 2003). This assumption is not always true in the climate system, especially considering different forcing scenarios and spatial heterogeneity of projected change. For temperature-related variables the assumption of stationarity is generally valid, but the magnitudes of estimation errors vary between scenarios for non-temperature variables (*Frieler et al.*, 2012) and temperature extremes on the upper tail of the temperature distribution (*Lustenberger et al.*, 2014). *Lopez et al.* (2013) found that when pattern scaling temperature extremes over Southern Europe, the magnitude of the error in the pattern estimates was substantially large. In linear regression, only the error term (ε) is assumed to have a normal distribution (based on the central limit theorem), so it is highly likely that climate extremes would yield high error 20 terms. This can be problematic when constructing confidence intervals but is not necessarily a limitation in the pattern scaling methodology, nor in the resulting patterns (*Lustenberger et al.*, 2014).

---

8. Later it is stated that the delta method to construct a pattern assumes that the trend within the (e.g.) 30-year period is the same regardless of epoch (or forcing scenario). No, it doesn't. See your equation (1): the delta method is simply the ratio of two mean differences, and means over epochs can be computed regardless of whether there is a trend during the epoch. The conclusions make a further related claim, with no support: "the delta method assumes that there is no observed trend in the historical simulation". Where does such an idea come from? If true, it would invalidate the delta method in almost all cases, since nearly all GCMs simulate a warming trend over the historical period.

Response:
We agree that the stating that the statement "…regardless of epoch chosen, the trend is the same" is incorrect. The assumption of GMT change and local change being independent of trend is important especially when considering the length of the epoch. The idea is that a longer epoch will increase the signal-to-noise ratio, by decreasing variance, but may also decrease the signal. This idea is first explored by Mitchell, 2003, and later by Barnes and Barnes, 2015. This assumption in no way invalidates pattern scaling, and we regret that our statements were interpreted to imply that it does. However, in this manuscript we do not explore how epoch lengths may change the resulting pattern. It remains unclear how epoch length affects the resulting pattern.
While we believe this to be an issue that warrants future study, this discussion does not belong in the introduction. We have edited this paragraph and moved a portion of it to the methods section where we justify our use of the 30-yr epoch length.
Also, we agree that the statement "…the delta method assumes that there is not observed.." is incorrect, and this sentence has been edited.

Revised text:

In results section: "When using the delta method to construct a pattern, the assumption is that regardless of epoch chosen, the trend is independent of change."

In conclusion section: "The delta method introduces further complexity in choice of reference epoch and length of reference epoch."

---

9. The "Assumptions" section then correctly states that the regression method assumes that local changes scale proportionally to GMT and that this relationship is stationary over time. This is correct, but the implication is that this is particular to the regression method – it is not, it applies equally to the delta method and indeed it is really an assumption of the PS approach itself and not additionally an assumption of the method used to diagnose the PS pattern. This is followed up by two unjustified statements: "Transient forcing is likely to scale the local temperature sensitivity to the trend in global mean temperature" (doesn't make sense and is not generally true) and "For temperature-related variables the assumption of stationarity is valid" (not necessarily so, e.g. local temperature change over areas of sea-ice retreat or snow cover retreat).

Response:
Yes, we agree that our statement implied that stationarity is unique to the regression method. We have corrected this statement in the manuscript to state that stationarity is assumed across all pattern scaling methodologies.
The statement "Transient forcing is ….." was indeed confusing, and does not add value to the paragraph, but it was meant as a way to imply that local temperature change scales very well with GMT change. However, it is not needed, and this statement has been removed.
The statement "For temperature-related variables…" should not imply that for all regions stationarity is a valid assumption. This statement has been edited, and a discussion of areas where stationarity is not valid for temperature is given in the results section.

---

10. The manuscript then rather hopefully claims that "the differences between the two methods are clear" despite the confusion sown by the errors detailed above – and indeed it is rather a forlorn hope since there aren't any major differences in the assumptions that underlie these pattern diagnosis methods. The underlying assumptions are those of the PS approach itself, which apply equally to both pattern diagnosis methods. The differences will arise because deviations from these assumptions will affect different time periods, simulations and calculations differently and thus the diagnosed patterns will depend on these choices. If these are systematic effects that can usefully guide best practise, that would be interesting – but the manuscript says nothing about this. Further differences arise because some methods have a stronger signal-to-noise ratio than others – but again the manuscript says nothing about this.

Response:
We agree that the introduction had not made clear the fundamental differences between the pattern methodologies. We have spent considerable time editing and rearranging the introduction based on comments from all three reviewers.
We also briefly examine the issue of signal to noise by using the methodology in Hawkins and Sutton, 2012, but this is only comparable to the regression method. For the delta method, the 30yr time period doesn't incorporate large variability, and spatially, the difference in variance across epochs is small and unlikely to affect the signal in a substantial way. We have included a figure in the Supplementary Material that shows this.

11a. The work needs to distinguish assumptions and errors in pattern diagnosis from assumptions and errors in applying the PS approach to generating future projections. This requires consideration of how to assess PS performance, which is lacking. Yes, there are differences between patterns, but which give the closest emulation of the GCM simulation?

Response:
Yes, we agree that assessing PS performance has been difficult. We have attempted to do that by examining the differences between patterns and scenarios, and explaining where and why those differences exist. Estimating emulator fit for the regression method has been done with the addition of a figure that shows the r^2 between global and local temperature. However, without using a scaler from our SCM, PS performance estimation is limited. We believe that in lieu of a scaler, Figure 8 most correctly estimates performance.

11b. Significance tests are claimed to show where the linear fit is poor, but the test does not discriminate between poor linear fit and a good linear fit for a weak relationship (e.g. a region where there is little warming in the GCM simulation may have an insignificant relationship with GMT). The first EOF in global annual temperature may give the warming trend, but this does not require GMT to have a linear trend, and anyway it is not necessarily so (this would depend on the strength of the forced signal relative to the internal variability among other things) so this statement is incorrect and it is unclear what the purpose of the EOF analysis is.

Response:
What we had hoped to show with the EOF table and supplementary figure is that the warming signal is clear. This ties into the signal-to-noise ratio by showing the percentage of the variance explained by the warming signal across historical and future scenarios, as well as showing where the signal is the strongest.

11c. The pattern differences and scenario differences sections (which are rather brief, but presumably are the main purposes of the paper) make a number of comparisons and find a number of differences, but it is unclear what is being shown and what the interpretation is. For example, some high latitude differences are put down to Arctic amplification – but this is an inadequate explanation, since this is in the GCMs and can be captured by PS if it is linearly related to GMT (just with a coefficient > 1).

Response:
We have expounded on both the pattern differences and scenario difference sections and included more thorough discussion of results. We agree that our brief statement about Arctic amplification is insufficient. We have included a more detailed explanation of high-latitude fit issues and their causes in the discussion of the results.

11d. The comparison with the GCM output is unclear and possibly not independent: comparing the PS trend patterns to the GCM trend patterns for the same simulations from which the PS pattern was diagnosed will give an overly optimistic view of the performance of the PS regression method. If the purpose is to establish whether PS based on delta patterns performs better or worse than PS based on regression patterns at emulating GCM projections, then more thought needs to be given to how performance is measured.

Response:

We agree that the regression method produces a very similar result to GCM estimated changes. We acknowledge that the assumptions of stationarity and linearity are very apparent when using the regression method and that the comparison with the GCM output is possibly not independent. We have further discussed this figure in the results section.

---

12. There are some further issues (e.g. Fig. 1a is inconsistent with Fig 9.8 of IPCC AR5 WGI, which gives 1961-1990 observed GMT, from Jones et al., 1999, around 14 °C and GCMs scattered around it – whereas Fig. 1a here has annual mean GMT around 10-11 °C for 1961-1990) (e.g. confused use of "variability" – unclear if it means temporal climate variability or ensemble variability/inter-model spread) but of diminishing importance.

Response:

We have moved this figure to supplementary, as the absolute values are not important in favor of anomalies, and it does not add to the points we make in the manuscript. However, after review of the AR5 WG1 Figure 9.8 we have found that the method used to construct GMT for the CMIP5 models applies a HadCRUT4 observational data mask. This would likely cause the resulting GMT at 1961-1990 to be higher than our GMT timeseries because large portions of the Arctic and Antarctic are not included in the GMT calculation in Figure 9.8. The section where variability is discussed has been edited, due to inaccurate assessment of variability in the original manuscript. We have added an additional figure in the main text as well as a figure in the supplementary material to further address this issue. Additionally, the idea of signal to noise ratio has been examined briefly to discuss the role of variance in pattern creation.
