Seventh Framework Programme

Theme 6

Environment

Project: 603864 – HELIX

Full project title:

High-**E**nd c**L**imate **I**mpacts and e**X**tremes

Deliverable: 2.4

Assessment of regional climate change & effectiveness of pattern scaling approaches

Version 1.0

Original Due date of deliverable: 30/04/15

Actual date of submission: 29/04/15

Assessment of Regional Climate Change and Effectiveness of Pattern Scaling Approaches

April

2015

Assessment of regional climate change and effectiveness of pattern scaling approaches

Contents

HELIX DELIVERABLE 2.4:

Assessment of Regional Climate Change and Effectiveness of Pattern Scaling Approaches

Craig J. Wallace, Timothy J. Osborn

Climatic Research Unit, School of Environmental Sciences, University of East Anglia

Jason A. Lowe, Dan Bernie

United Kingdom Meteorological Office Hadley Centre, Exeter

28 April 2015

EXECUTIVE SUMMARY

Pattern scaling is a widely-used method for creating multiple climate change scenarios with which to investigate climate change impacts. Although generally sufficient to capture the climate-change response of General Circulation Models (GCMs) over land, the technique has known limitations and its apparent accuracy is also sensitive to the validation data set to which it is compared (especially the ensemble size).

For this deliverable we examine and quantify the effects upon performance of the 'ClimGen' pattern scaler in emulating a specific GCM (HadgEM2-ES under RCP8.5 forcing) arising from both the validation data ensemble size and potential non-linearities in the climate change response of the GCM. We especially investigate the emission-dependence of the pattern-scaler – that is investigating whether patterns diagnosed from GCMs under some emission scenarios are acceptable for use in pattern scaling to emulate other (namely high-end) scenarios.

We show here that sensitivity of performance to the ensemble size of the validation data set is large, as use of multiple ensemble members (limited to 4 in this work) strengthens the true climate change signal which the pattern scaler is attempting to reproduce. Even accounting for this fact, however, we identify degradation in the pattern-scaling performance when mean global warming approaches 3.5°C. This can be attributed to a tendency for patterns to underestimate the GCM warming over many land areas. Of note to the HELIX community is the relative superior performance of the pattern diagnosed from the high-end (RCP8.5) GCM simulations for generating pattern-scaled projections for global warming of more than 3.5°C in comparison to other patterns including ClimGen's default pattern which would typically be used in generating scenario data.

Re-calculating performance metrics whilst spatially limiting the pattern-scaled and validation data to exclude regions where HadGEM2-ES is known to exhibit non-linear climate responses shows little affect upon apparent performance. This is unexpected, especially for global warming levels of 6°C. The seemingly small influence may be a consequence of our limited data pool for high-end warming (just one ensemble member for validation) whereas lower warming thresholds lie within the scope of the full ensemble of four members.

1 Introduction and Pattern Scaling Overview

The pattern-scaling (PS) approach enables the generation of gridded, time-varying, climate change scenarios combining the spatial climate-change responses of a number of general circulation models or Earth system models (collectively GCMs) with a driving time series of global-mean temperature change. The driving time series of global temperature changes (hereafter *deltaglobalT*) and the patterns of GCM climate-change responses (hereafter *patterns*) can be diagnosed from any future climate change scenario (e.g. the IPCC representative concentration pathways, RCPs) for which the GCM itself has been forced with in a conventional transient manner, providing the climate change signal is large enough (compared to the GCM's unforced variability). Alternative global temperature changes can be prescribed, either fixed specific warming levels (SWLs) or transient changes simulated by simple energy-balance climate models that can explore a wider range of future scenarios and model uncertainty than is typically possible for GCMs. In effect, pattern scaling is an approximate emulator for the more complex GCM behaviour in terms of its geographical, seasonal and multi-variate response to global anthropogenic forcing. Thus, the appeal of PS is that a wide pool of climate change scenarios, including GCM-forcing combinations not included in the training data, can be quickly generated which encompass both GCM-related and scenario-related uncertainties. The approach is popular within the context of coupled physical and socio-economic environmental prediction frameworks (e.g. Arnell et al. 2014; CIAS: Warren et al. 2008; IMAGE: Van Vuuren et al. 2006) and climate impact studies (e.g. Gosling and Arnell 2013; Ostberg et al. 2013) where the coupling (and running) of an suite of GCMs is unfeasible.

Although computationally attractive, the PS approach is only an approximate emulation of conventional transient GCM simulations. This is principally because pattern scaling (described in more detail below) encapsulates climate change responses in a linear form, or with simple first-order functions. The real response of a GCM (and the real climate system) has some features that do not evolve so simply (e.g. Good et al.,2015). Early investigation of PS methods (Mitchell et al. 1999; Mitchell 2003) revealed such non-linearities but also established that the consequent pattern-scaling errors were small compared to the size of the uncertainties relating to other aspects of the climate prediction (e.g. forcing selection; GCM choice; the climate signal itself). Recent analyses have continued to assess the PS method and found it to be sufficient to approximate the greenhouse gas responses of the latest generation of GCMs, such as CMIP3 (Henke et al. 2013) and CMIP5 (Tebaldi and Arblaster 2014), but performs less well where strong regional differences in forcing exists (e.g. sulfate aerosols: Ishizaki et al. 2012) and for climate variables with upper and/or lower bounds (e.g. cloud amounts; precipitation). The PS tool considered in our analysis here (*ClimGen*, below) can account for some aspects of the latter, and so for this HELIX deliverable we first provide some general assessment of the technique (Section 2) using the latest simulations of the HadGEM2-ES GCM (ref) before focussing upon two sources of error that influence PS validation and which are of relevance to the application of PS data within the context of the HELIX framework – especially to making climate projections under the specific warming levels (SWLs) of 2 °C, 4 °C and 6 °C:

1: The first focus (Section 3) examines the affect of validating PS-generated climate scenarios against either a single transient GCM run (of the GCM being emulated by the PS approach) compared with validation against the mean of an initial-condition ensemble, in which the forced climate change

signal is implicitly clearer. This exercise enables attribution of the apparent PS errors to either the GCM-realization-specific unforced variability or to the genuine limitations of the PS approach in capturing the forced climate change response.

2: The second focus, building on the first, examines the proportion of the PS error which can be ascribed to known features of non-linearity (Good et al. 2015) in the climate change response of the HadGEM2-ES climate model and how the size of this error differs between varying combinations of GCM-diagnosed patterns and local, grid-cell temperature responses.

1.1 The *ClimGen* pattern scaler

In a basic form, a future change in a climate variable, *V*, at a spatial grid cell at some time in the future, *t*, can be estimated by:

$$\boldsymbol{V_t} = \boldsymbol{a}\Delta T_t \quad (1)$$

where ΔT_t is the *deltaglobalT* at time *t*, relative to a pre-identified baseline (e.g. 1961–1990), diagnosed in this case from a transient run of a particular GCM under some future greenhouse gas concentration profile. and $\boldsymbol{a}$ is the grid cell specific pattern coefficient for the variable in question representing the linear change in the variable per degree of global-mean warming. The pattern coefficients can be diagnosed from a single transient run of the GCM under a specific climate change scenario (e.g. RCP2.6) or may be the diagnosed simultaneously from several runs of the same GCM, but each under a different forcing scenario (e.g. RCP2.6 and RCP4.5 data pooled together). An important question is whether the pattern-scaling coefficient is a function of emissions scenario or climate pathway.

The *ClimGen* pattern scaler (Osborn et al. 2015) applies the approach of (1), using fields of $\boldsymbol{a}$ for each of the twelve calendar months, to calculate air temperature change with a seasonal cycle. To diagnose the spatial coefficients, $\boldsymbol{a}$ is calculated via a linear regression of each land grid cell's anomaly through the full GCM simulation against the accompanying *deltaglobalT*, with appropriate time-filtering to remove interannual variability. Thus, within the pattern scaling approach used here, $\boldsymbol{a}$ is assumed to be independent of time. *ClimGen* is also able to add interannual variability to the resulting scenario data, but this option is not considered for the purposes of the validation analyses presented here, focussing instead on changes in mean climate. For precipitation, a more complex approach to (1) is used in *ClimGen* to represent different GCM behaviour under drying or wetting conditions and changes in the shape of the distribution of interannual variability). The validation analyses here are limited to near-surface air temperature change over land, though the findings have implications for other variables over land.

The native spatial resolution of *ClimGen* is 0.5° latitude by 0.5° longitude (land only) and so patterns of $\boldsymbol{a}$ are spatially interrpolated from the GCM's native grid to the 0.5° grid after they have been diagnosed. For this HELIX deliverable, however, we validate *ClimGen*-calculated data against the actual GCM transient climate data for the **HadGEM2-ES** GCM and so we leave the spatial resolution of the pattern-scaler on the native GCM grid and undertake all comparisons on this grid.

We also note that in many studies pattern-scaling and bias correction of GCM data are both applied together. This report is only focusing on the pattern-scaling aspects. Bias correction will be investigated elsewhere in HELIX.

1.2 Patterns and GCM Data used for the validation and pattern-scaling projections.

Throughout this analyses we use the pattern-scaler method to attempt to reconstruct the transient GCM air temperature of **HadGEM2-ES** under the RCP8.5 greenhouse gas concentration profile. The validation data set is composed of two parts:

i: a single-member extended run under RCP8.5 forcing running **2001 - 2299**, with CO_2 concentrations stabilising by 2250 at 2000ppm.

ii: an additional three simulations, again under RCP8.5 forcing, but covering **2001 – 2100** only. Forcing here has not yet stabilized (Jones et al., 2011).

Validation of the pattern-scaled data against a high-end forcing scenario (RCP8.5) is a more robust test than, say, moderate scenarios with lower warming, since nonlinear deviations from the linear pattern-scaling approximation may well scale with forcing, and so will become more apparent.

For the pattern-scaling calculation, $\Delta \boldsymbol{T_t}$ is the global-mean monthly air temperature for 1951-2299 derived from the historical-forcing HadGEM2-ES simulation and both parts of the validation data described above (thus it is an ensemble mean up to 2100 and a single member thereafter). These data are expressed as anomalies from the simulated 1961-1990 mean and smoothed using a 10-year Guassian-weighted filter followed by a spline (Fig. 1). The purpose of this smoothing is to reduce unforced interannual variability in the future, but (via a less-stiff spline) to retain forced climate features within the historical phase of the transient runs that vary on decadal scales (e.g. from volcanic forcing). Using the 1961-1990 period as the reference baseline for the pattern-scaling means resulting climate scenario data (anomalies) are obviously benchmarked to this period also. For the validation data set (the HadGEM2-ES transient data) we also anomalize the GCM data to this period . Using the 1961-1990 period for the 'present' day is the default within *ClimGen* hence its use here, although an alternative reference period is often the pre-industrial phase, typically defined as 1861-1890. The difference in *deltaglobalT* between the two mean 30-year mean periods is approximately 0.35°C (according to the HadCRUT4 temperature record) and so using different reference periods will have some affect on the illustrative years (in the pattern-scaled and direct GCM projections) selected to represent the SWL threshold timings. This affect is small, however, amounting to ~7 years difference in each case of the 2C, 4C and 6 C SWLs threshold timings, with differences in actual anomalies also reduced by the time filtering implicit in the PS data and an analogous filtering applied to the validation HadGEM2-ES data (applied to ensure statistical congruence between the PS and validation data).

The pattern coefficients, $\boldsymbol{a}$, are diagnosed from HadGEM2-ES data also, but from the 1951–2100 period only and for a variety of forcing permutations (Table 1). Pattern-scaled climate scenarios can then be calculated by combining each of the pattern-permutations in Table 1 with $\Delta \boldsymbol{T_t}$. Validating a number of pattern-permutations will isolate the impact of climate change patterns that are a

function of emissions scenario and/or climate pathway. This is an important aspect of the validation, since pattern-scaling applications often apply pattern coefficients calculated from one RCP simulation to the $\Delta \boldsymbol{T}$ from another, with no acknowledgement of the pattern dependence due to the actual response of the GCM under each RCP.

Pattern name:	Transient HadGEM2-ES data (1951-2100) used to calculate pattern
RCPall	RCP2.6, RCP4.5, RCP6.0, RCP8.5 (four ensemble members in each)
RCP264560	RCP2.6, RCP4.5, RCP6.0 (four ensemble members in each)
RCP26	RCP2.6 only (four ensemble members)
RCP85	RCP8.5 only (four ensemble members)

Table 1: The various combinations of RCP-forced transient GCM data used to generate spatial pattern coefficients for the pattern scaling projections.

The HELIX programme is concerned with climate change at SWLs of 2 °C, 4 °C and 6 °C global-mean temperature rise. Therefore we select years at various time points in the projected data sets where *deltaglobalT* reaches these SWL thresholds (Fig. 1) and examine the performance of each pattern-scaling permutation against the analogous time period of the GCM validation data set (a 30-year mean, centred on the year that the SWL is reached).

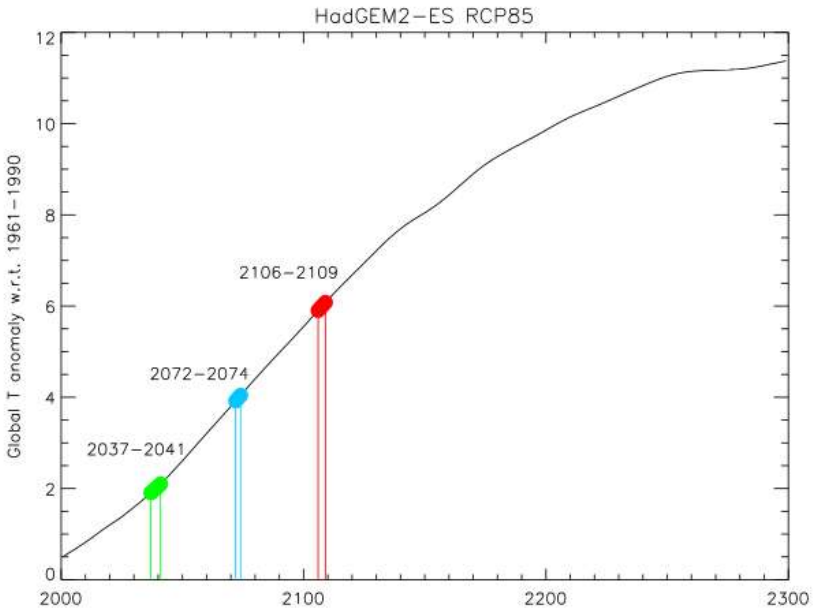


Fig 1: HadGEM2-ES deltaglobalT (RCP8.5 monthly) used in the pattern scaling calculation. Green, blue and red locations show years where SWLs of 2, 4 and 6 °C, relative to 1961–1990, *are approximately reached (+/-0.2deg). Sample years used for these SWLs in this analysis are* ***2038*** *(****2*** °C*),* ***2073*** *(****4*** °C*) and* ***2108*** *(****6*** °C *). Using smoothed data increases the likelihood that reaching the SWL threshold arises from the climate change signal, rather than realization-dependent unforced variability.*

2: Validation of Pattern scaling performance

Having generated pattern-scaled climate projections (hereafter *PS*) using each of the permutaions in Table 1, for all months covering 2001-2299, we can examine performance via error metrics when the PS data are compared against the validation transient HadGEM2-ES data set (hereafter *TRANS*) . We consider some general perfomance overviews where TRANS is the single-member HadGEM2-ES data (i.e. 'i' in section 1.3) before examining, in more detail, the influence of **ensemble-member variability** and **climate system non-linearity** (see 1.1) on the performance of the pattern scaler. To ensure consistency of the TRANS and PS data, all gridded TRANs data are smoothed using a 30-year running mean to dampen interannual, unforced variaility that is not the goal of PS**.**

Patterns of the PS-TRANS differences for representative 2, 4 and 6 °C SWL years highlight areas where the PS technique underestimates or overestimates the warming simulated by HadGEM2-ES (Fig. 2). Broadly speaking a general tendency for a cool bias in the PS air temperature can be seen, although the sign of errors is not ubiquitous (e.g. there is a warm bias over Eurasia at 4 °C SWL in January). In fact, the cool bias appears to be balanced, if not dominated, by warm biases for winter (January) for 2108 in the projection series (equating to SWL 6 °C). This tendency towards warm bias under increased warming is not apparent in the summer (July).

To place the magnitudes of the signed PS-TRANS errors into context, we compare them with the magnitude of the actual TRANS projected air temperature change (i.e. the HadGEM2-ES climate signal the PS is attempting to reproduce). This can be done by computing the root mean squared error (RMSE) of all land grid cell (PS-TRANS) values, for each month through time, together with the root of the mean squared TRANS climate anomaly again for each land grid cell and, in all cases, area weighted). When compared as such (Fig. 3) it is clear that the PS errors are a small fraction of the climate change signal. The ratio, in fact, decreases from ~0.3 to ~0.15 in all months (not shown) as global warming approaches 3-4°C increases, after which the gradually ratio rises to around 0.2 as *deltaglobalT* surpasses 10°C. The sigmoid type shape to the anomaly signal when plotted through time (Fig. 3b) can be attributed to the stabilization of the RCP8.5 forcing mid way through the scenario. Tebaldi and Arblaster (2014) confirmed the overall validity of the pattern-scaling approach as an approximate representation of the CMIP3 and CMIP5 multi-model ensembles. This "validity" arises because the errors in pattern scaling were shown to be small relative to the ensemble spread.

For example, their Fig. 4 compares the spread of climate change patterns from different models and the spread of climate change patterns under different RCP scenarios; the latter, which represents non-stationary or non-linear behaviour that PS cannot always capture, is shown to be only a small portion of the total variation. This applies to both temperature and precipitation projections. We find similar results in our own analysis of the CMIP5 ensemble. For annual temperature, around 10% of the local variance across the ensemble arises from differences in the normalised patterns of change between scenarios for the same model. There is some spatial variation, but in only a few locations does the contribution rise above 20%. For annual precipitation, the average contribution of scenario-dependent patterns is greater (~20%), though unforced variability is also greater for precipitation and tends to cause an overestimate of the estimated contribution

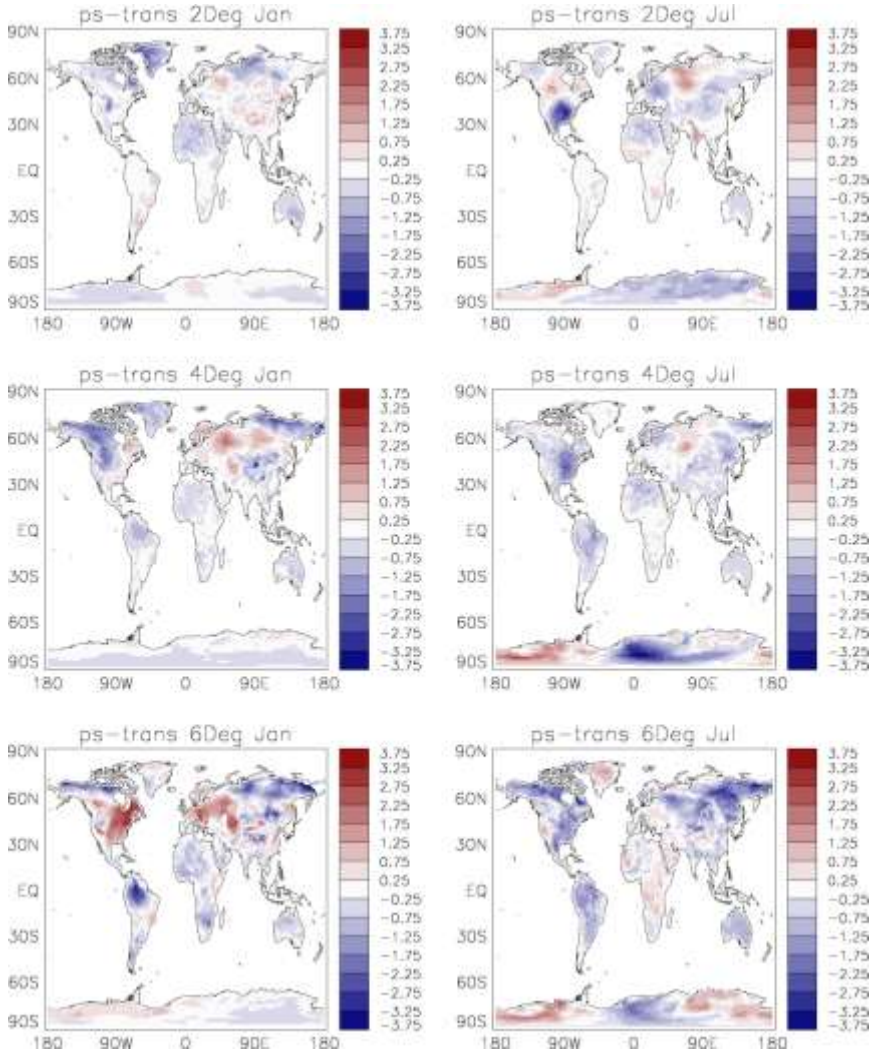


Fig 2: Grid cell *PS–TRANS* (°C) for January (left) and July (right) for years when *globaldelta*T = 2 °C (2038 top); 4 °C (2073 middle); 6 °C (2108 bottom). PS projections are calculated using the *RCPall* pattern, which is the default pattern selection in the *ClimGen* pattern-scaler.

The pattern used to produce the PS projections in Fig. 3, namely *RCPall*, was diagnosed using all HadGEM2-ES simulated data out to 2100. The comparison against HadGEM2-ES RCP8.5 beyond 2100 (or beyond ~5.5 °C of global warming) is, therefore, an independent test of the PS approach, while the comparison up to 2100 includes data that were used to diagnose the patterns being tested. A further test can be made using a pattern diagnosed from all simulations except RCP8.5 (namely *RCP264560*), which is then independent of the testing data throughout (green line in Fig. 4). It is also informative to compare the relative accuracy of patterns diagnosed from individual GCM simulations by repeating Fig. 3 but for each permutation's RMSE (Fig. 4).

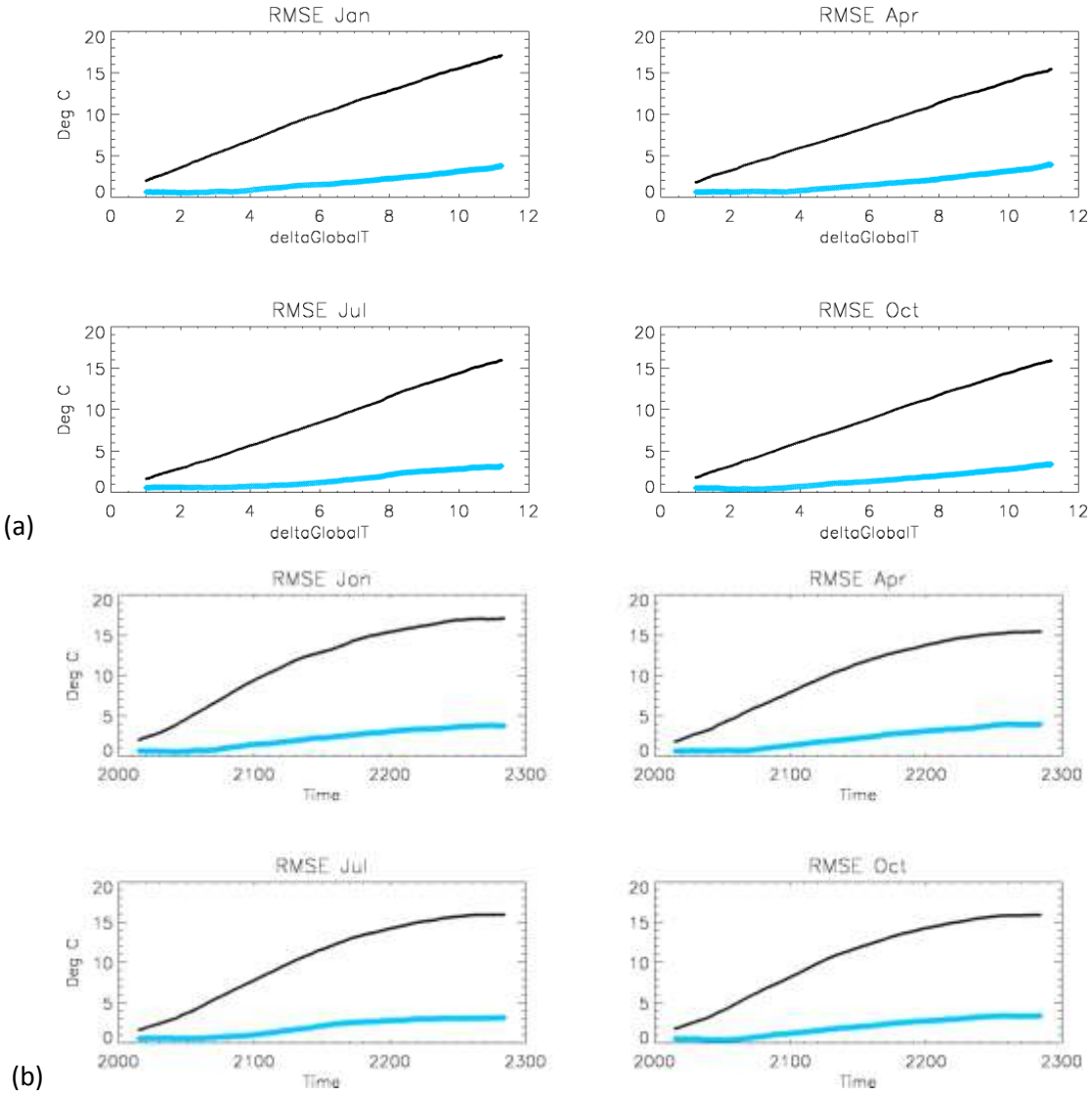


Fig. 3: Global RMSE of gridded *RCPall* PS air temperature compared against 1 ensemble-member TRANS (blue) together with the global RMS of TRANS climate anomaly (**black,** from same ensemble member) for RCP8.5. Series shown as function of *deltaglobalT* (a) used for the pattern-scaling calculation and time (b).

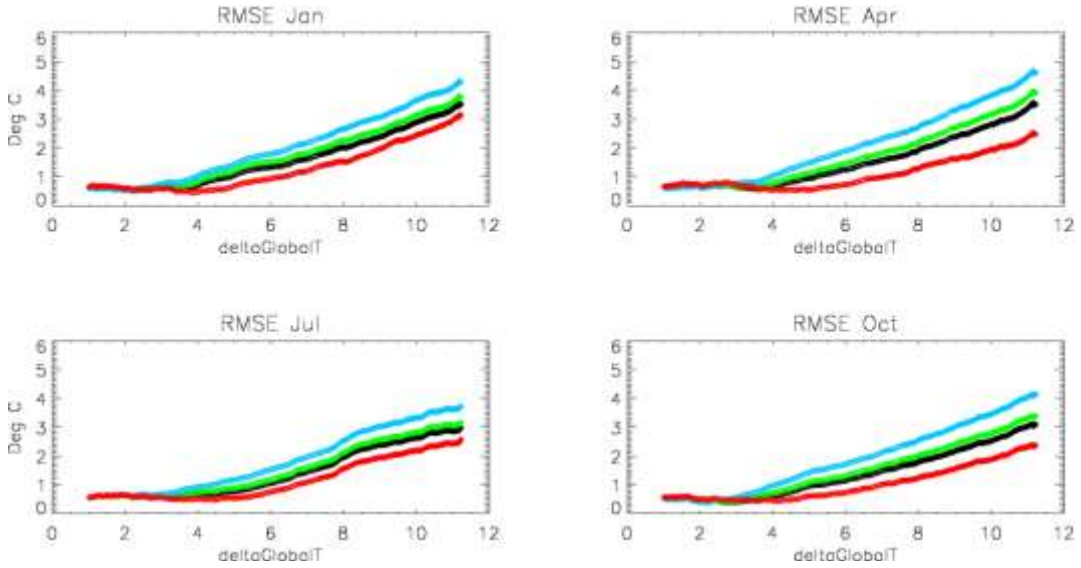


Fig. 4. Global RMSE of PS temperature for variety of patterns (***RCPall***; ***RCP264560***; ***RCP26***; ***RCP85***) compared against HadGEM2-ES gridded projections from one ensemble member (2001 – 2299) under RCP8.5. RMSE are expressed as a function of the RCP8.5 global warming (i.e. *deltaglobalT* used for the pattern-scaling calculation).

Examining the additional RMSEs, as a function of the scaling *deltaglobalT*, shows the superiority of the *RCP8.5*-only $\boldsymbol{\alpha}$ (red curve, Fig. 4) over the other patterns for reproducing the GCM projection under RCP8.5. *RCP26* performs worst, which we can attribute to the pattern-coefficients, $\boldsymbol{\alpha}$, having been diagnosed in a low-forcing situation and then extrapolated to higher temperatures. Using patterns from the RCP2.6 simulation but scaled by RCP8.5 *deltaglobalT* is attempting to represent an increasingly large forcing behaviour, likely to have a clearer response than that of RCP26's $\boldsymbol{\alpha}$ and possibly non-linear behaviour that may not be evident in the relatively low forcing levels of the RCP2.6 training data**.** It is interesting to note the divergence of errors beginning around the **3 °C** *deltaglobalT* threshold (approximately ~2050), which must be linked to such factors, whereas RMSE preceding this threshold are low and similar in all PS cases.

Looking at the spatial patterns of errors for the best and worst performing patterns (*RCP26* and *RCP85*) near this point of performance divergence can tell us more about the nature of the performance disparities (Fig. 5). Both PS permutations have a tendency, in places, to underestimate the HadGEM2-ES transient warming at this time, but the cold bias is much stronger in *RCP26*. It is possible that non-linear vegetation and/or moisture-driven interactions lead to greater warming responses, for example over the American continent, that are encapsulated in the *RCP85* patterns (since sufficient climate forcing is present in the data from which $\boldsymbol{\alpha}$ is derived) whilst they are missing from the RCP26 patterns of coefficients (derived from a much lower forcing situation in which these mechanisms may not have emerged). It seems that with globaldeltaT > 3 °C such mechanisms emerge in the TRANS data (RCP8.5 HadGEM2-ES), leading to a greater error with

respect to *RCP26* PS data than *RCP85*. Similarly, the RCP26 overestimation of the RCP8.5 temperature response in eastern Canada could be related to ocean dynamics present in both the *RCP85* pattern diagnosis data and the TRANS data (by year 2073 at least), but not the *RCP26* pattern diagnosis data.

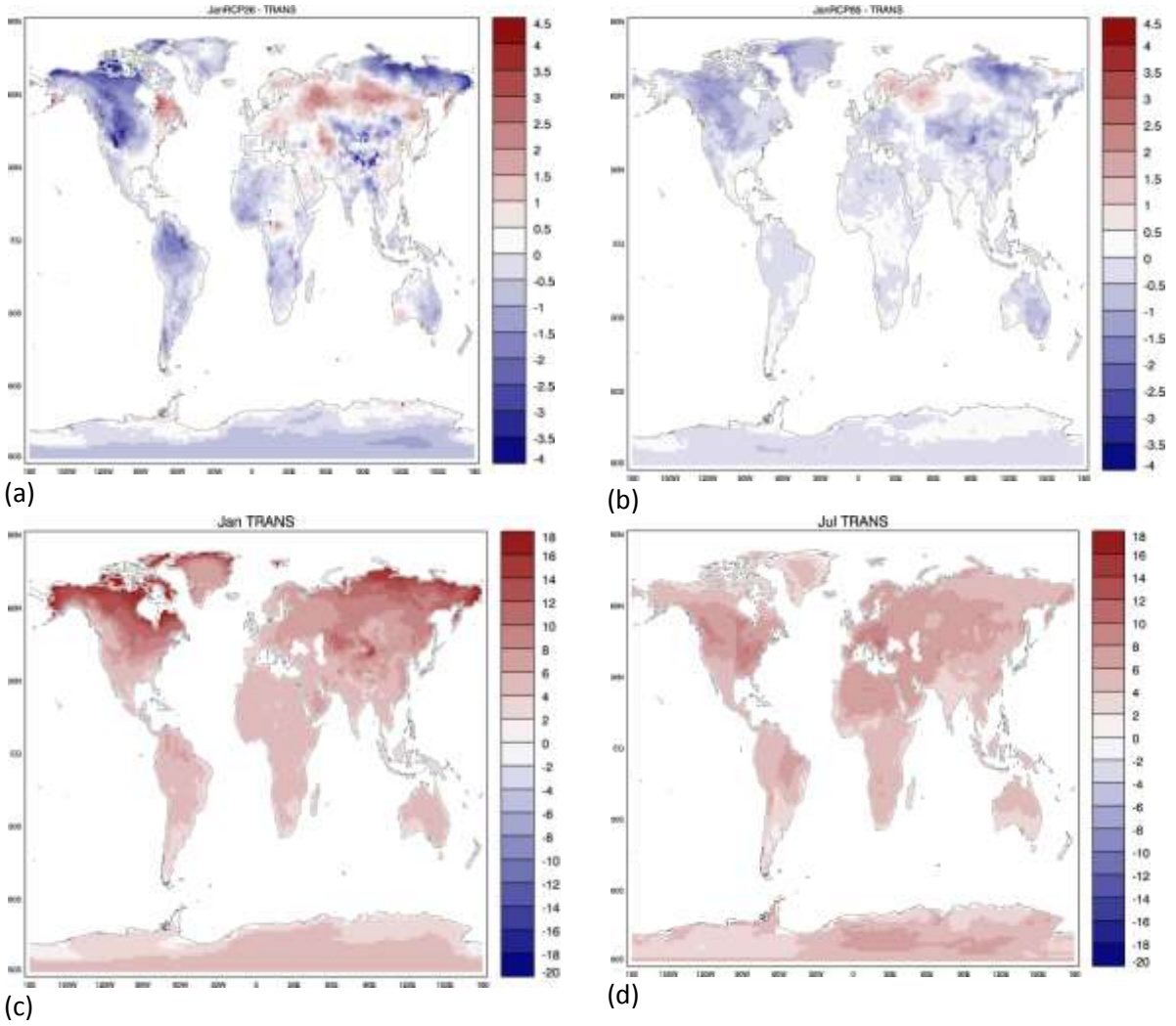


Fig. 5. Maps of pattern scaling errors (°C) for the *RCP26* (a) and *RCP85* (b) patterns when compared against the HadGEM2-ES transient data for January in year 2073, when global warming = 4°C. Panels (c) and (d) shows the spatial pattern of climate change in Jan (c) and Jul (d) for the same year, from the validation HadGEM2-ES transient data.

Comparing PS data with TRANS, when TRANS is a single simulation of HadGEM2-ES, is useful (as in Fig2, 3, 4), but is not a perfect measure of the pattern-scaler's ability to reproduce the bona-fide climate change signal of the GCM because an individual realization of any GCM will have a unique, naturally-driven climate evolution component, unassociated with external forcings such as greenhouse gases. The RMSEs considered so far, then, will be a combination of the pattern-scaler's deficiencies in fully emulating the GCM forced climate change response and the unforced variability simulated by the GCM. The comparisons have been made using 30-year means, so unforced

variability on timescales shorter than 30 years will not have inflated the RMSE values, only unforced variability on timescales longer than 30 years.

We can quantify each of these contributions to the overall RMSE by recalculating PS-TRANS RMSE but using the mean of all four RCP8.5 ensemble members (i and ii in 1.3) as the target climate projection, TRANS, to dampen the unforced variability present in the GCM simulations. This comparison is limited to the 2001-2100 period for which the ensemble is available. Recalculation of the RMSEs shows marked reductions (Fig. 6). In fact, for global warming up to ~3 °C, errors are approximately halved. As greenhouse warming proceeds, the climate change signal grows in comparison to the unforced variability and the reduction in errors degrades: beyond ~3 °C this becomes visible in a steepening of the ensemble-mean PS-TRANS curve (red).

Given that the unforced variability in each ensemble member is independent, the standard deviation of unforced variability at each grid cell will be reduced in an ensemble mean by the square root of the ensemble size. With a 4-member ensemble, we would expect the RMS of the unforced variability to have been halved compared with a single ensemble member, and that, if we had an infinite ensemble of runs to reduce the unforced variability to zero, its RMS would be reduced by the same amount as the reduction already seen in going from one to four ensemble members. In other words, the 4-member ensemble mean still contains significant levels of unforced variability and this unfairly degrades the apparent performance of PS at capturing the forced climate change signal in tests such as those presented in Fig. 6. Indeed, were the reduction in RMSE from the single run results (black lines) to the 4-member results (red lines) doubled, to represent an infinite ensemble with no unforced variability, it would appear that the RMSE of the pattern scaled projections would be close to zero for global warming up to 3 °C under RCP8.5. However, this is an over-optimistic interpretation because the PS-based projection is not a perfect representation of even the linear component of forced climate response due to some contamination of the patterns by residual unforced variability (although the patterns are diagnosed by pooling data from all ensemble members across all RCP scenarios, and by regression of temporally-smoothed changes against simulated global-mean warming, to reduce this contamination). Nevertheless, these results do confirm that, when evaluated against the GCM's climate change signal rather than a combination of signal and unforced variability, the errors arising from the PS technique are small compared with the projected climate changes themselves.

Further investigation into the contribution of ensemble-member variability to PS errors is shown in **section 3.**

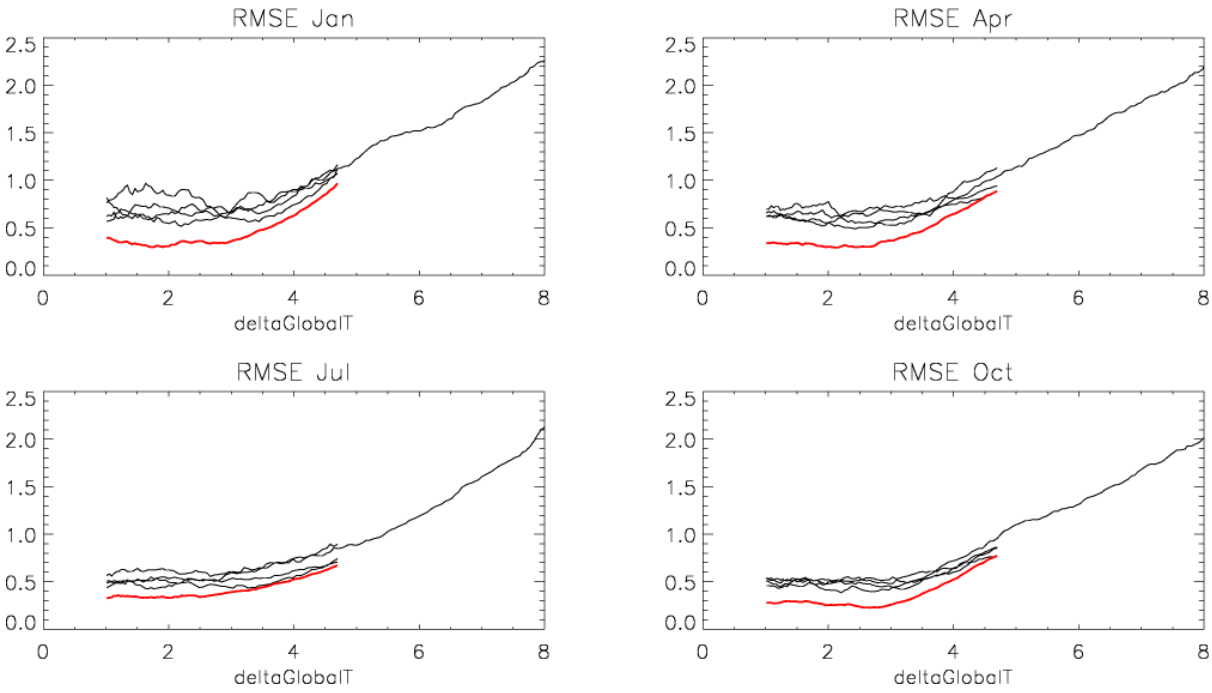


Fig. 6: Global RMSE of PS land surface air temperature projections (*RCP264560*) when compared against gridded projections from each of the four HadGEM2-ES ensemble members (**black** lines – one for each ensemble member; only one member covers the extended climate change period). **Red** is RMSE but calculated against the HadGEM2-ES ensemble *mean* projections, for available 2001-2100. RMSEs are a function of the *deltaglobalT* used in the pattern scaling calculation (x axis).

It is also useful to examine the performance of each PS permutation as a function of the climate-change anomaly in each grid cell. The grid cell climate-change anomaly (*localT*) is taken as the surface air temperature anomaly in each cell from the validation HadGEM-ES RCP8.5 simulation (TRANS). For a sample projection year (e.g. 4 °C global warming, the 30-year mean centred on 2037) all land grid cell errors (i.e. the PS projection minus the single HadGEM2-ES RCP8.5 simulation covering 2001-2299) can be ordered according to increasing *localT*, then accumulated and displayed as a function of *localT* (Fig. 7). This allows us to explore and association between the size of the PS error and the strength of the local warming. The final height of each curve in Fig. 6 is the total accumulation of errors across each spatial field of projections, thus ranking the PS permutations from best-to-worse as *RCP85*, *RCPall*, *RCP264560* and *RCP26* (consistent with Fig. 4). The steepness of each curve indicates at what localT each permutation accumulates the greatest errors. The steepness is also related to the clustering of grid cell counts, since the error will accumulate more quickly the more grid cells there with similar local temperature change. This is depicted by the vertical dot-dashed lines, showing the position of the five quintiles of local temperature change and demonstrating the concentration of changes around 4-5 °C *localT* in January (Fig. 7a).

In January, for instance, *RCP26* and *RCP264560* gain errors relatively more rapidly than *RCP85* and *RCPall* above 1 °C *localT*, while *RCPall* errors are similar to RCP85 errors until ~4.5 °C *localT* before increasing more rapidly. Thus the poorer performance of *RCPall*, relative to *RCP85*, arises in regions that warm by 5 °C or more. We can attribute more than half of the overall error to grid cells in the

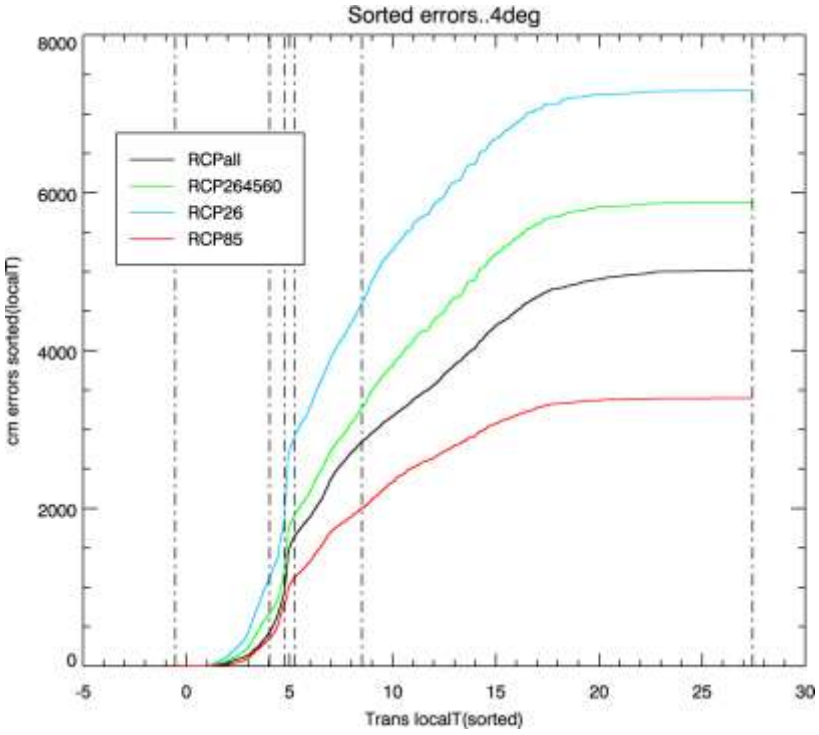
upper two quintiles of local warming (Fig. 7a), consistent with a degeneration of performance with local warming from ~5 °C and above – as shown in the spatial pattern of errors for the illustrative year (2037) Fig 5a. Beyond 5 °C *localT*, the error accumulation of all permutations except *RCP85* is very large, demonstrating a decrease in PS performance for other permutations at larger delta *localTs*.

Examining accumulated errors for July (Fig. 7b) we see that the spread of *localT* is narrower than in January (maximum *localT* = 10.5 °C) but the overall accumulated errors are in the same order and reach approximately the same level as for January. In July, however, the contribution of errors from cells in the uppermost quintile of warming is not so great as it was in January, with greater errors instead occurring in some regions with more moderate warming, consistent with the spatial patterns of the PS errors (Fig. 5c) and local TRANS warming (Fig. 5d). Another notable feature of Fig. 7b is that the *RCPall* pattern does a better job at projecting the 4 °C pattern under RCP8.5 than the *RCP8.5* pattern in regions where warming is between 3 and 4.5 °C (i.e. where the black line is less steep than the red line).

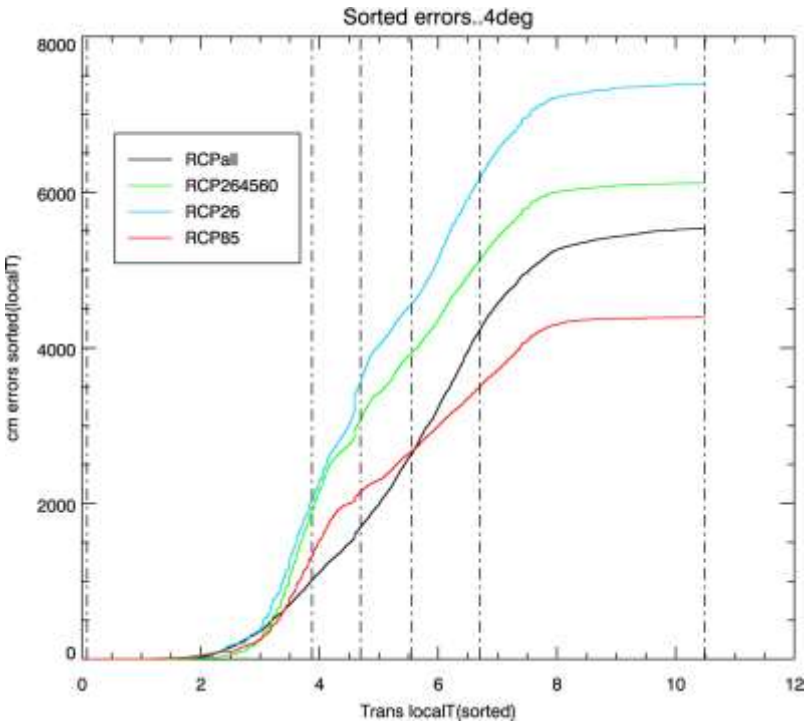
3. Examining contribution of GCM unforced variability to pattern scaling evaluation in more detail

Here, we further examine the improvement of the PS errors, for each permutation, which arises when we evaluate the PS projections against the 4-member HadGEM2-ES ensemble mean rather than against a single GCM simulation (Fig. 6). We can also display these improvements as a function of grid-cell climate anomaly, *localT* (as in Fig. 7, but now *localT* is the ensemble-mean grid-cell temperature change). To do so, for each illustrative year of the projections associated with SWL, we re-calculate grid cell absolute errors (as in Fig. 7), but using the difference between the PS-based projection and the 4-member HadGEM2-ES ensemble mean. This new error, in each cell, we call ***Eens***, as opposed to ***Esgl*** when it is the difference from the single HadGEM2-ES simulation**.** The comparison of Eens and Esgl is limited to 2 °C and 4 °C SWLs, since global-mean temperatures do not reach 6 °C (above 1961–1990) until ~2108 and the 4-member ensemble ends in 2100.

Having calculated *Eens* and *Esgl* at each grid cell for each PS permutation, we express them as a ratio (*Eens*/*Esgl*) at each *localT* to quantify the fraction of error related to the higher unforced GCM variability in the single simulation, through *localT* space. However, *localT* at each grid cell will be different in the *Eens* and *Esgl* populations, so we first bin the errors by their *localT*, using a nominal 0.1 °C bin size, and averaged *Eens* and *Esgl* errors in each bin. The bin means can then be expressed as (*Eens*/*Esgl*) for all grid cells in an illustrative SWL year of the projections (e.g. Fig. 8 for global warming = 4 °C). Eens/Esgl < 1 indicates error reductions when comparing with the GCM ensemble mean. We also show in Fig. 8 the density of grid cell counts in each *localT* for both the ensemble and single-member HadGEM2-ES data. For the case of a winter month (January, Fig. 8a), we see consistent reductions in errors coinciding with the largest grid-cell densities at *localT* ≈ 3.5-6 °C, especially for *RCPall* and *RCP85* PS permutations. The improvements for *RCP26* (and to some extent for *RCP264560*) are less clear, suggesting in this case, the errors are more dominated by real differences in the climate change signal between PS and TRANS, with a smaller contribution from unforced variability in the GCM simulations.



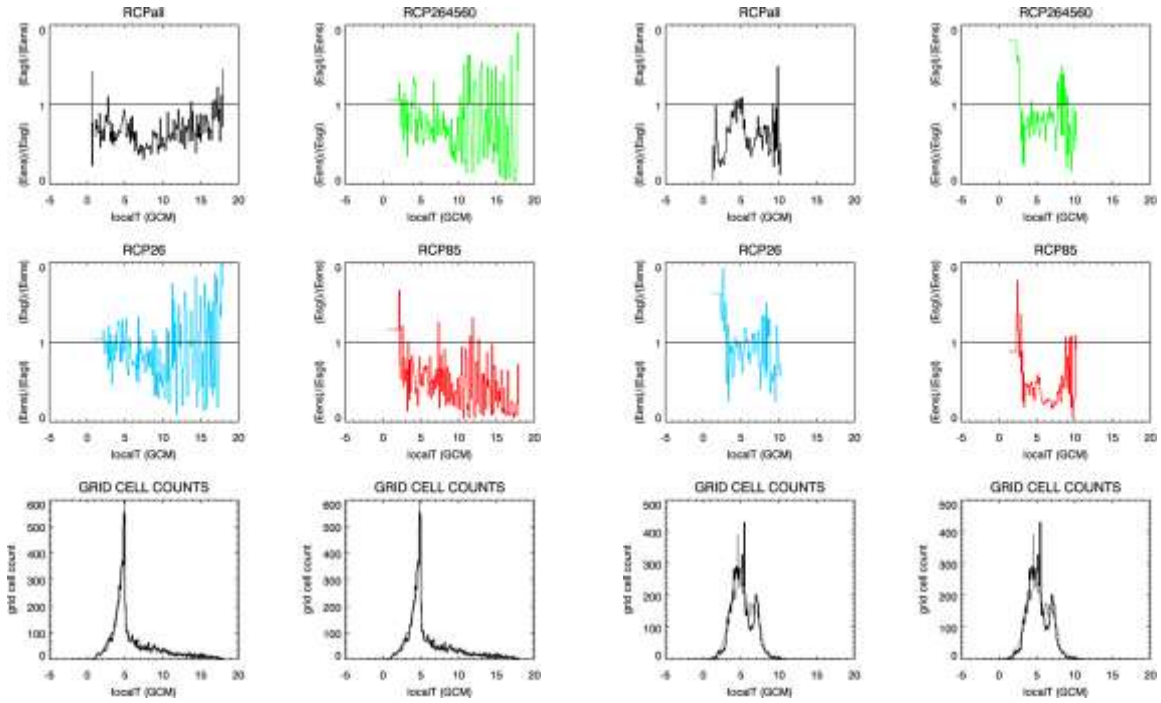
(a) Jan



(b) Jul

Fig. 7: Accumulated grid cell (PS-TRANS) absolute errors (y-axis, °C) as a function of grid cell *localT* anomaly (x axis, °C) for year 2073 of the projection for (a) January and (b) July. Each coloured line represents errors for each of the PS permutations used for the pattern-scaling calculation. Vertical dashed lines indicate quintiles of land grid cell counts (through *localT*) within the validation data set (TRANS).

For July (Fig. 8b) we can see a wider peak density of cell counts and fewer grid cells with higher *localT* changes. This distribution is consistent with the strongest greenhouse warming in NH winter months and with the comparative quintile spacings in Fig. 7a verus Fig. 7b. In the case of *RCP85*, error reductions are again evident (i.e. *Eens*/*Esgl* < 1), although for *RCPall* and *RCP264560* the error changes are perhaps less consistent than comparative January values The *RCP26* permutation remains fairly poor, as in the case of January. It is useful to recall that the apparent 'improvement' in the PS performance in Fig. 8 are a result of the clarification of the real climate change signal in the *validation* data, rather than any refinement of the PS projections (or technique).



(a) (b)

Fig. 8: *Eens*/*Esgl* fractions (y axis) for PS permutations **RCPall (black)**, **RCP264560 (green)**, **RCP26 (blue)** and **RCP85 (red)** in Jan (a) and Jul (b) as a function of grid cell localT (0.1 °C bins, x axis) for year 2073 (with 4 °C global warming, relative to 1961–1990) of the PS and TRANS projections. Bottom duplicated panel shows grid cells counts for each localT bin for the single-member TRANS data (thin) and ensemble-mean TRANS data (thick). Note, where *Esgl* > *Eens* (and therefore *Eens*/*Esgl* > 1), the numerator and denominator are switched, as is the direction of the y axis.

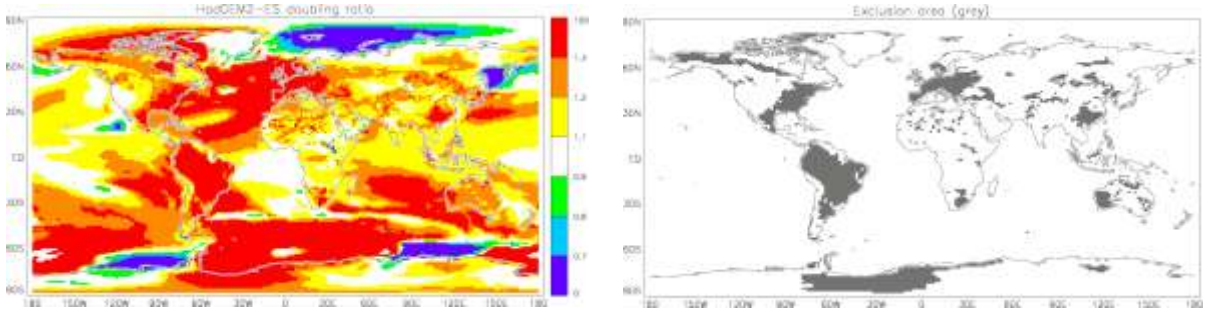
4. Climate system non-linearity as a source of pattern-scaling errors

A limitation of the pattern-scaling approach is the assumption of a linear response within a region or grid cell to each successive increment in global warming. Moreover, although the ClimGen pattern scaler accounts for the different behaviour of some climate variables (e.g. precipitation, which can follow an exponential evolution where it reduces towards zero, or cloud cover, which follows a

logistic function), the assumption remains that the diagnosed per-degree response is stationary. In reality, aspects of the climate system are known to behave non-linearly, responding in a different manner to sequential forcings which are of identical sizes (e.g. the response of Atlantic Ocean circulation). Such processes influence different regions of the land surface temperature. Recent analyses (Good et al. 2015) compared the changes in air temperature after two successive doublings of CO_2 simulated by HadGEM2-ES to derive a linearity metric, the ***doubling ratio***, according to:

$$\frac{\Delta T_{db2}}{\Delta T_{db1}} \tag{2}$$

where ΔT_{db1} is the change in a grid cell temperature from a control climate to a 2xCO_2 state, and ΔT_{db2} is the further change from the doubled state to a 4xCO_2 state. The HadGEM2-ES pattern of doubling ratio (after Good et al. 2015) is shown in Fig. 9a. Deep red indicates grid cells where warming after the second doubling is greater than after the first; blue/purple depicts the reverse. Red Atlantic and European sectors can be attributed to the non-linear weakening of the Atlantic Ocean overturning and follow-on effects to regional air temperature. High-latitude features with doubling ratios below one are related to changes in snow or sea ice – rapid under the first doubling, but then stabilising (thus eliminating feedbacks) in the second. Zones of high doubling ratios over South America can be attributed to a combination of vegetation, precipitation and soil moisture dynamics (Good et al. 2015).



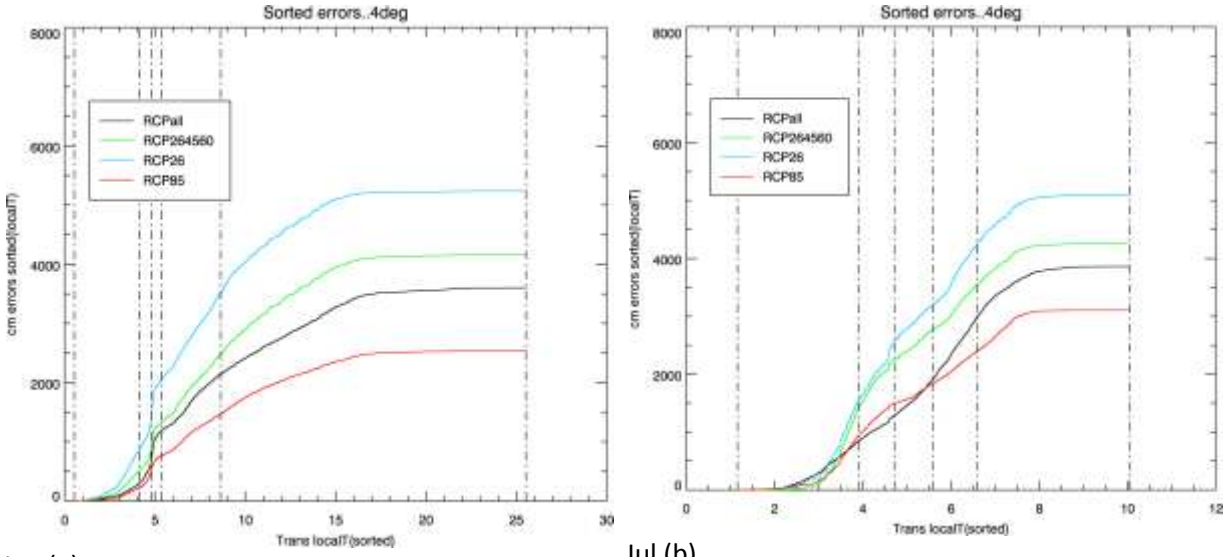
a b

Fig. 9a: the doubling ratio from Good et al., 2015 9b: the land grid cells where the doubling ratio is less than 0.75 or greater than 1.25 (grey). These grid cells are excluded to leave a nominal area where the climate response is linear.

To quantify the magnitude of the effects of these known HadGEM2-ES non-linearities on the PS errors, we repeat the *Eens* and *Esgl* error analysis (Section 3) but this time comparing the full spatial field '***Full***', and a limited spatial field (excluding areas of high or low doubling ratios) '***Lin***'. To proceed, we generate a spatial mask (Fig. 9b) based upon the HadGEM2-ES doubling ratio values in which we exclude grid cells where the doubling ratio is <0.75 or >1.25 (an arbitrary selection), leaving only grid cells with a low level of non-linearity in their temperature response. It is important to state that the thresholds of the doubling ratio used to define the mask will likely affect the

magnitude of error differences. Nonetheless, using the mask as a template upon the pre-calculated (PS-TRANS) grid cell errors, for each SWL, we can then have a population of spatially-limited errors (***Elin***) to complement the existing full field errors (***Efull***). Note, the PS-TRANS errors in both calculated with respect to the single-member HadGEM2-ES simulation.

Compiling the masked population of errors, *Elin*, for the 4 °C SWL (represented by year 2073) enables us to plot the accumulated errors as a function of *localT*, as for the full field in Fig. 7. Comparing Fig. 7 and Fig. 10 shows a similar quintile distribution of grid-cell counts in both January and July, however the error accumulation (both the final amount, and at various points in *localT*) is obviously affected by the smaller number of grid cells in *Lin*. Thus it is difficult to used Fig. 10 to attribute the obvious lower final error amounts to either the simple exclusion of grid cells (26% are excluded compare to the full field) irrespective of error values, or to the exclusion of error cells on that have higher-than-typical errors because of non-linearities.

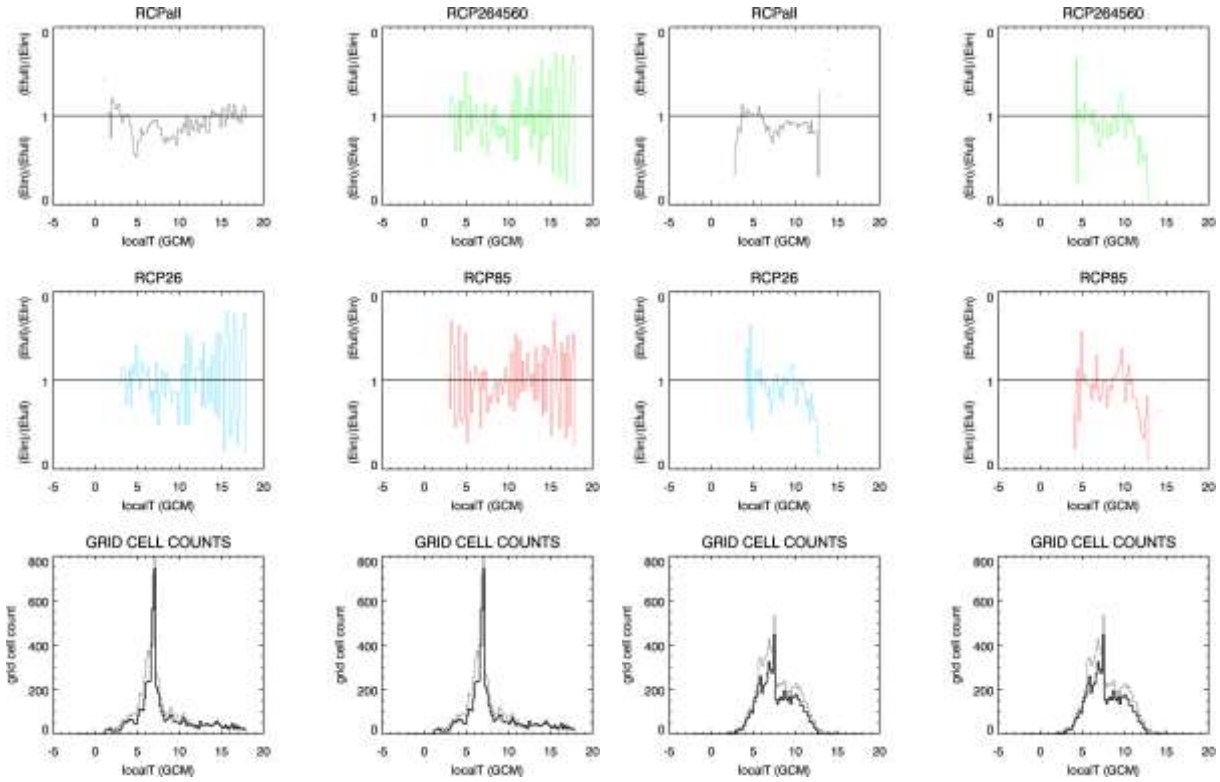


Jan (a) Jul (b)

Fig. 10: Accumulated absolute (PS-TRANS) errors (y-axis, °C) for grid cells within the masked field as a function of grid cell *localT* anomaly (x axis, °C) for year 2073 of the RCP8.5 projection. The masked field excludes the grid cells with the most nonlinear behaviour according to the doubling ratio of Good et al. (2015). Jan (a); July (b). Each coloured line represents errors for each of the PS permutations used for the pattern-scaling calculation.

A better method to quantify non-linearity effects, using the *Lin* mask, and accounting for changing grid cell counts, is to average grid-cell errors for *Full* and *Lin* populations over 0.1 °C bins of grid-cell *localT* and to calculate ratios of these mean errors for each bin. As with the *Eens* versus *Esgl* error comparison (Fig. 8), *Elin*/*Efull* can then be plotted to depict the fractional change in PS error resulting from the exclusion of 'non-linear' cells. Some changes are evident when this metric is calculated for *deltaglobalT*=4 °C (not shown) although larger changes are evident by year 2106 (when *deltaglobalT* = 6 °C) when forcing may trigger the non-linear behaviours found by Good et al. (2015). The resulting plot (Fig. 11) shows apparent reductions in the errors for the most heavily sampled *localT*

bins (i.e. ratios < 1, coinciding with spikes in grid cell count spikes) for *RCPall*. Improvements for the remaining PS permutations are less clear, although there is some consistent reduction in RCP264560 and RCP85 errors for July. It is perhaps at first surprising that the average errors in RCP26 do not show more of an improvement. This seems to suggest that non-linear climate responses (at least of the nature defined by Good et al. 2015) do not greatly influence the PS performance at deltaglobalT = 6 °C. For the RCP26 pattern, diagnosed from a HadGEM2-ES simulation where deltaglobalT is only ~1.7 °C in 2100, it may simply be that the extrapolation to 6 °C warming is dominated by errors across the field and not only in those identified by the Good et al. doubling ratio. One final consideration is that there may be additional non-linearities outside the doubling ratio mask.



Jan (a) Jul (b)

Fig. 11: As Fig. 8, but for the ratio *Elin*/*Efull*, comparing the errors in the full fields with those masked to exclude the grid cells with the most nonlinear behaviour according to the doubling ratio of Good et al. (2015), and for an example future year where *deltaglobalT* = 6 °C (2106) for the single HadGEM2-ES ensemble member which runs 1859-2299. Bottom panels show grid cell counts in each 0.1 °C bin for *Lin* (thick) and *Full* (thinner) populations.

5. Summary

We have investigated the performance of the popular pattern-scaling technique for emulating the climate change projections from one particular GCM, HadGEM2-ES. We focussed on climate changes over land, where the linear assumptions that underlie the pattern-scaling approach are more reasonable and where most pattern-scaling applications are used. We evaluated changes in near-surface air temperature, though the approach is applicable to other variables. Two factors which influence the validation performance of pattern-scaling were considered further. First, the impact of unforced climate variability present in the GCM simulations used for validation, and second, the impact of non-linearities in the climate change response of the HadGEM2-ES model. We have presented results relevant to specific warming levels of interest to the HELIX community – principally global-mean warming of 4 and 6 °C.

It is possible to identify errors in the PS projections, which is expected since this is only an approximation to the complex, non-linear behaviour of GCMs. However, the errors are relatively small compared with the size of the climate change signal (Fig. 3), and also compared with other sources of uncertainty in climate change projections, namely the choice of forcing scenario and the spread amongst an ensemble of multiple climate models.

Furthermore, evaluations of PS projections against GCM-simulated climate changes are hampered because the GCM simulations contain unforced natural variability that the PS technique is not intended to emulate. Using the mean of an initial-condition ensemble of GCM runs reduces the magnitude of the unforced variability and notably improves the similarity between the PS projections and the direct GCM results (Fig. 6; note this apparent 'improvement' in the PS performance results entirely from the clarification of the real GCM climate change signal in the *validation* data, rather than any refinement of the PS projections or technique).

In Fig. 12, we extend the results of Fig. 6 (which were only for the case where the pattern used for scaling had been diagnosed from pooled HadGEM2-ES simulations under RCP2.6, RCP4.5 and RCP6.0 i.e. *RCP264560*) to show the results using different patterns diagnosed separately from only RCP2.6 or RCP8.5 simulations, or from pooling all RCP simulations including RCP8.5. The latter is of particular interest because this is the default pattern used in the ClimGen pattern-scaler. In all cases, the test is to emulate the HadGEM2-ES land surface air temperature during 2001–2100 under RCP8.5, comparing with the GCM's 4-member ensemble mean. All patterns result in projections with RMSE around 0.5 °C or less for most cases up to about 3.5 °C of global warming. Beyond this, the patterns that we diagnosed without using the RCP8.5 simulation get progressively worse (in fact the RCP2.6 only pattern diverges earlier), while the pattern from the RCP8.5 simulation which performed least well for global warming below 2 °C has gradually reducing errors, reducing a minimum around 4 °C of global warming. Fig. 4, though not evaluated against an ensemble mean, suggests that the RCP85 pattern maintains its advantage out to 6 °C and beyond, and t*his should be considered in the generation of climate scenarios for impact work under high-end scenarios.*

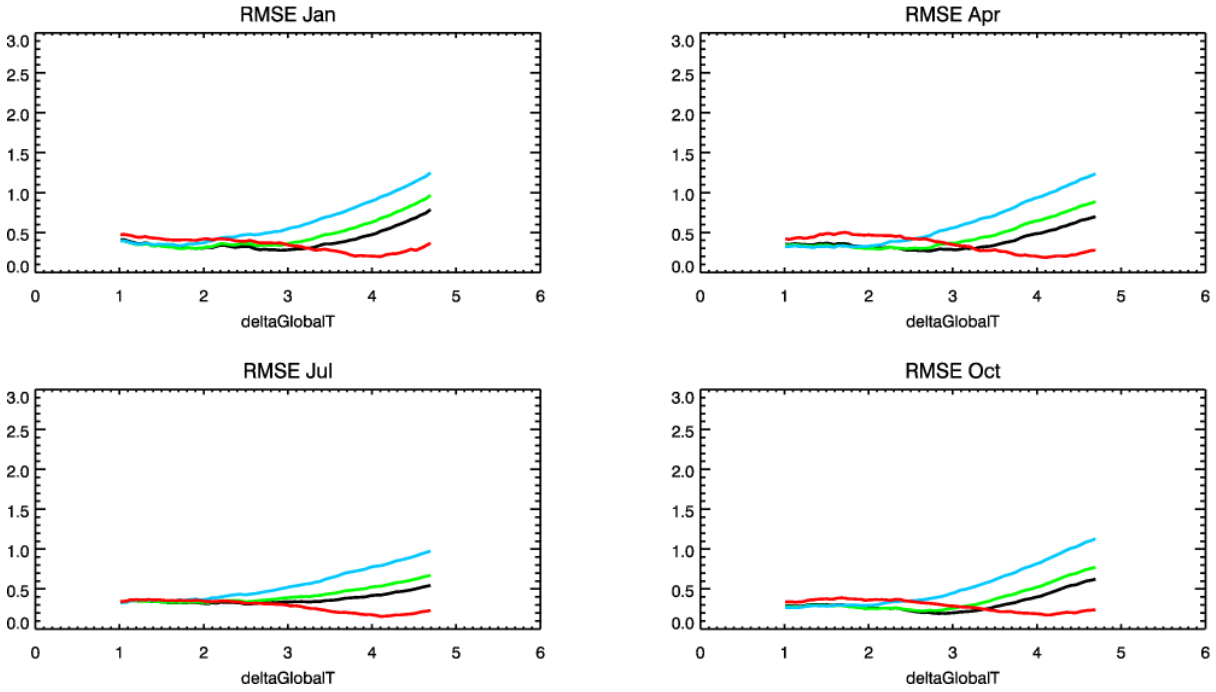
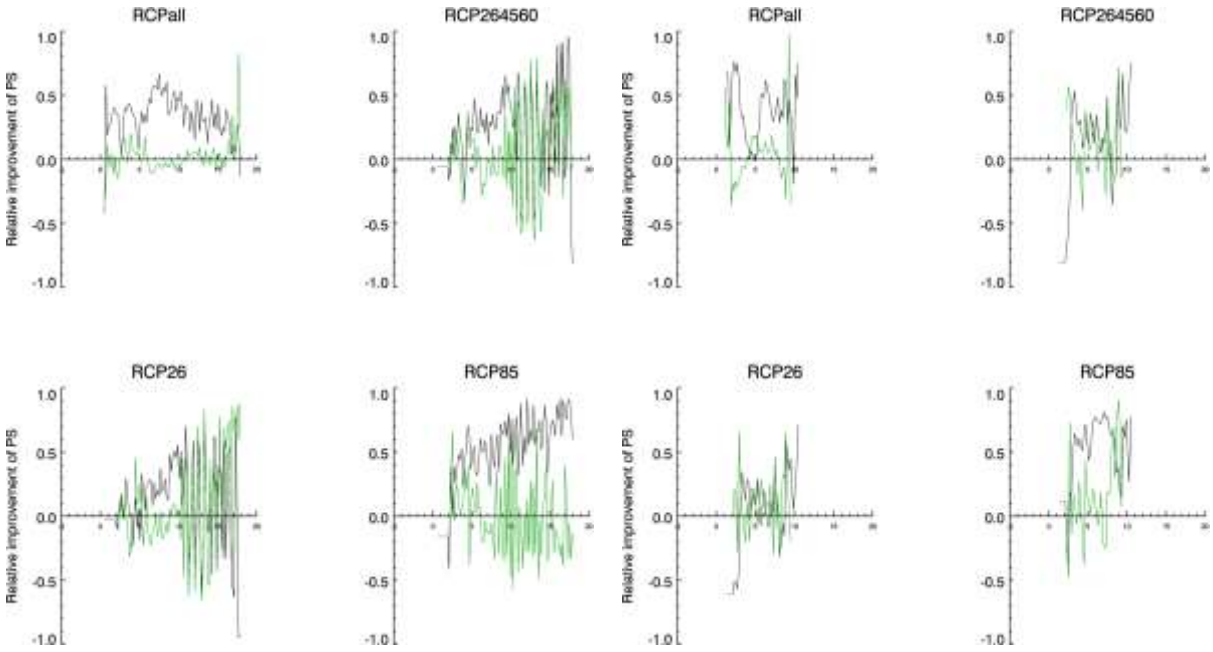


Fig. 12. Global RMSE (degC) of PS permutations **RCPall (black)**; **RCP264560 (green)**; **RCP26 (blue)** and **RCP85 (red)** when tested against the ensemble-mean HadGEM2-ES grid cell land surface air temperature projections under RCP8.5 from 2001–2100, plotted against the global-mean temperature change during the simulation.

Even with a 4-member ensemble mean, and evaluating climate changes using a 30-year time-mean, there will still be significant unforced variability in the spatial fields of surface air temperature change (and even more so for other variables such as precipitation). This means that even the performance metrics based on comparison with the GCM ensemble mean will be conservative – i.e. they overestimate the pattern-scaling errors.

With respect to general performance of the PS technique (e.g. Figs. 4 and 12), we have shown the growth of errors for global warming above 3 to 4 °C, for all patterns other than *RCP85*, is related to an underestimation of the transient HadGEM2-ES RCP8.5 warming (Fig. 5) in some regions (especially with the *RCP26* pattern). Fig 8 quantifies these errors and their improvement when validating against the 4-member ensemble mean as a function of the size of the local climate change. The errors have also been investigated with respect to known regions of non-linearity in the HadGEM2-ES model. Figs 10 and 11 depict the size and character of this potential influence according to which pattern is used to make the RCP8.5 climate projection, showing some small or moderate influences upon grid cell errors. It is important to note that for this analysis the errors both for the full and masked spatial fields have been calculated using only one GCM ensemble member to enable the masking exercise to be done beyond 2100 to reach a global warming of 6 °C. Results such as those in Fig. 11 are illustrative of the impact of non-linearities, but some of this may be attributable to the natural unforced variability unique to the single ensemble member being used. Nonetheless we can compare the relative size of the error adjustments resulting from (i) the linearity mask (from Fig. 12), and (ii) the ensemble-size exercise (from Fig. 8), for each pattern being

considered, by plotting fractional improvement (or worsening) of errors as a function of local grid-cell temperature change. Such a comparison is shown in Fig. 13, with values above (below) zero indicate improvement (worsening) of the error arising from the better validation using the 4-member ensemble (black) or masking out regions with known non-linear behaviour in the HadGEM2-ES GCM (green). This is only possible for the 4 °C SWL, since the ensemble size exercise cannot be done for the 6 °C SWL. What is clear in many cases is the relative size of the errors related to ensemble-size (unforced variability) is larger than those related to the non-linear behaviour, despite expecting the latter to be larger. Further work will repeat the quantification of the masking effects, but using the ensemble-mean HadGEM2-ES data as the RMSE validation data.



(a) (b)

Fig. 13. Magnitude of improvement in the error of pattern-scaled land surface air temperature projections for RCP8.5 (with respect to the HadGEM2-ES transient data) arising from validating against a 4-member HadGEM2-ES ensemble rather than a single GCM run (black) and masking out regions with known non-linear temperature responses (green). In latter, errors are evaluated by comparison to only a single ensemble member of HadGEM2-ES ensemble. (a) January and (b) July for year 2073 when global temperature rise reaches 4 °C above the 1961–1990 mean, as a function of (x-axis) the magnitude of the local grid cell temperature change simulated by the GCM.

References

Arnell, N. W., Brown, S., Gosling, S. N., Hinkel, J., Huntingford, C., Lloyd-Hughes, B., Lowe, J. A., Osborn, T., Nicholls, R. J. and Zelazowski, P. (2014) 'Global-scale climate impact functions: the relationship between climate forcing and impact', *Climatic Change*, doi: 10.1007/s10584-013-1034-7

Arnell, N. W., Lowe, J. A., Brown, S., Gosling, S. N., Gottschalk, P., Hinkel, J., Lloyd-Hughes, B., Nicholls, R. J., Osborn, T. J., Osborne, T. M., Rose, G. A., Smith, P. and Warren, R. F. (2013) 'A global assessment of the effects of climate policy on the impacts of climate change', *Nature Climate Change*, 3(5), pp. 512–519. doi: 10.1038/nclimate1793

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., C., Sitch, S., Totterdell, I., Wiltshire, A. and Woodward, S. (2011) 'Development and evaluation of an Earth-system model – HadGEM2', *Geoscientific Model Development Discussions*, 4(2), pp. 997–1062. doi: 10.5194/gmdd-4-997-2011

Gosling, S. N. and Arnell, N. W. (2013) 'A global assessment of the impact of climate change on water scarcity', *Climatic Change*, doi: 10.1007/s10584-013-0853-x

Heinke, J., Ostberg, S., Schaphoff, S., Frieler, K., Müller, C., Gerten, D., Meinshausen, M. and Lucht, W. (2013) 'A new climate dataset for systematic assessments of climate change impacts as a function of global warming', *Geoscientific Model Development*, 6(5), pp. 1689–1703. doi: 10.5194/gmd-6-1689-2013

Ishizaki, N. N., Shiogama, H., Takahashi, K., Emori, S., Dairaku, K., Kusaka, H., Nakaegawa, T. and Takayabu, I. (2012) 'An Attempt to Estimate of Probabilistic Regional Climate Analogue in a Warmer Japan', *Journal of the Meteorological Society of Japan. Ser. II*, 90B(0), pp. 65–74. doi: 10.2151/jmsj.2012-b05

Mitchell, T. (2003) 'Pattern scaling: an examination of the accuracy of the technique for describing future climates' *Climatic Change* ,60 ,pp. 217-242.

Ostberg, S., Lucht, W., Schaphoff, S. and Gerten, D. (2013) 'Critical impacts of global warming on land ecosystems', *Earth System Dynamics*, 4(2), pp. 347–357. doi: 10.5194/esd-4-347-2013

Tebaldi, C. and Arblaster, J. M. (2014) 'Pattern scaling: Its strengths and limitations, and an update on the latest model simulations' *Climatic Change*, doi: 10.1007/s10584-013-1032-9

Van Vuuren, D. P., Eickhout, B., Lucas, P. L. and Elzen, M. G. J. den (2006) 'Long-Term Multi-Gas Scenarios to Stabilise Radiative Forcing - Exploring Costs and Benefits Within an Integrated Assessment Framework', *The Energy Journal*, SI2006doi: 10.5547/issn0195-6574-ej-volsi2006-nosi3-10

Warren, R., S. de la Nava Santos, Arnell, N. W., Bane, M., Barker, T., Barton, C., Ford, R., Füssel, H.-M., Hankin, R. K. S., Klein, R., Linstead, C., Kohler, J., Mitchell, T. D., Osborn, T. J., Pan, H., Raper, S. C. B., Riley, G., Schellnhüber, H. J., Winne, S. and Anderson, D. (2008) 'Development and illustrative outputs of the Community Integrated Assessment System (CIAS), a multi-institutional modular integrated assessment approach for modelling climate change', *Environmental Modelling & Software*, 23(5), pp. 592–610. doi: 10.1016/j.envsoft.2007.09.002