Noname manuscript No.
(will be inserted by the editor)

Supplementary Information - Pattern scaling for decision support

Ana Lopez, Emma B. Suckling and Leonard A. Smith

Received: date / Accepted: date

1 Verifying the linearity assumption

The pattern scaling approach states that if $T(\mathbf{x}, t)$ is the actual pattern of change in the variable T at position $\mathbf{x}$ and year t simulated by a full GCM, an *approximate* pattern of change $T^*(\mathbf{x}, t)$ for this variable can be obtained in terms of a spatial pattern $P(\mathbf{x})$ and the global mean change $\hat{T}$ according to

$$T^*(\mathbf{x}, t) = P(\mathbf{x})\ \hat{T}(t), \tag{1}$$

where

$$P(\mathbf{x}) = \frac{\sum_t T(\mathbf{x}, t)\hat{T}(t)}{\sum_t \hat{T}^2(t)}, \tag{2}$$

is obtained as the spatial pattern that minimizes the distance $\sum_t [T(\mathbf{x}, t) - T^*(\mathbf{x}, t)]^2$ between T and T^* .

This approximation encapsulates the assumption that the spatial pattern of change $P(\mathbf{x})$ is constant in time, and the only effect of the transient forcing will be to scale the pattern up or down following the trajectory of the global mean temperature change. Hence "pattern scaling". The generalization of the above equation to include monthly or seasonal dependence is straightforward, the temperature change field $T(\mathbf{x}, t)$ becomes $T(\mathbf{x}, i, t)$ with i labeling a month or season in year t, and consequently there is a pattern $P(\mathbf{x}, i)$ for every possible value of i .

In this work the pattern $P(\mathbf{x})$ is derived using the I.C. ensemble mean, to improve the results with respect to scaling from an individual simulation [2]. The temperature change fields $T(\mathbf{x}, t)$ and global mean temperature change $\hat{T}(t)$ in equation (2) are those of the ensemble mean (represented by the thick black line in figure 1). The pattern derived in this way is named $P_{EM}(\mathbf{x})$, or $P_{EM}(\mathbf{x}, i)$ if the pattern is computed for each month or season.

Address(es) of author(s) should be given

If the linearity assumption is satisfied, the pattern derived using the ensemble mean time series, $P_{EM}(\mathbf{x})$, should be equivalent to a pattern obtained by considering the temperature fields of all the model runs when their global mean temperature changes coincide with each other, i.e., when they are all equal to some value T_G (which can of course take any arbitrary value). In order to test this assertion, and in analogy with equation (2), a pattern P_{T_G} is defined for any given global mean temperature change T_G as follows

$$P_{T_G}(\mathbf{x}) = \frac{\sum_{m=1}^{N} T_m(\mathbf{x}) T_m^G}{\sum_{m=1}^{N} (T_m^G)^2}, \tag{3}$$

where $T_m(\mathbf{x})$ is the temperature change field of model m when its own annual global mean temperature change is T_m^G, and N is the number of temperature fields. Notice that the number of temperature change fields is not necessarily the same as the number of model runs in the ensemble under consideration, since a model global mean temperature can be equal, or close, to T^G more than once due inter annual variability. In principle, we should set $T_m^G = T_G$ and the expression in equation (3) reduces to $P_{T_G}(\mathbf{x}) = \frac{\sum_{m=1}^{N} T_m(\mathbf{x})}{N T_G}$. In practice, due to the discreteness of the time series, T_m^G is not strictly equal to T_G and the sum in equation (3) is over the model temperature fields $T_m(\mathbf{x})$ every time its global mean temperature change is within a range of T_G that has to be predefined. Two cases are analyzed here:

Case I: choosing all T_m^Gs that are closer to T_G than the ensemble mean global mean temperatures right below and right above T_G ,

Case II: choosing all T_m^Gs that fall within a fixed interval around T_G, $T_G \pm 0.01$.

The generalization of equation (3) to a monthly or seasonal pattern is again straightforward, the temperature change field $T_m(\mathbf{x})$ becomes $T_m(\mathbf{x}, i)$ with i labeling the month or season of the year when $T_m^G = T_G$, and consequently the pattern becomes $P_{T_G}(\mathbf{x}, i)$.

It is clear that if the temperature field labeled by m satisfies the pattern scaling relationship defined in equation (1) for the pattern derived from the I.C. ensemble mean, i.e., if $T_m(\mathbf{x}) = P_{EM}(\mathbf{x})\ T_m^G$, then $P_{T_G}(\mathbf{x}) = P_{EM}(\mathbf{x})$. In other words, if the temperature field of any given model scales linearly with the global mean temperature changes, then the pattern obtained from the ensemble mean must be identical to the pattern obtained using the temperature change fields of the whole ensemble at any fixed global mean temperature change T_G. The three T_Gs analyzed in what follows, $T_G = 1.5°, 2.0°$ and $2.5°$, are indicated with horizontal colored lines in figure 1.

To compare the two patterns P_{T_G} and P_{EM} an estimate of the sampling noise must be provided. This is estimated by recalculating the value of each pattern by re sampling the corresponding ensemble as follows [1]:

Method I (leave one out): leaving one model out and computing the pattern using the remaining ensemble. In this case the number of samples is identical to the size of the ensemble minus one.

Method II (bootstrapping) : by re sampling the ensemble 1000 times with replacement.

Figure 2 shows the results for $P_{EM}(\mathbf{x}, i)$ (black cross) and $P_{T_G}(\mathbf{x}, i)$ (colored crosses) when $\mathbf{x}$ corresponds to Southern Europe and i corresponds to the annual average, and to two seasonal averages: boreal summer (JJA) and winter (DJF). Here $P_{T_G}(\mathbf{x}, i)$ is defined as in Case I. The error bars represent the 90% range of the re sampled ensemble using the methods I and II described above. For method I the confidence intervals are narrower: the ensemble mean changes slightly from sample to sample (only one model is left out at time and the ensemble size is always the same), therefore the estimated patterns are close to each other. For method II , the ensemble is potentially changing significantly from sample to sample (there will be repeated model runs in each re sampled ensemble), thus the variability in the estimated patterns is larger than in the previous case. In this case, if re sampling with replacement is interpreted as a proxy for sampling a very large I.C. ensemble, the error bars represent an estimate of the natural variability sampling uncertainty at the chosen confidence level.

As illustrated in figures 3 and 4 for Central North America and Southern South America, the results are region and time average period dependent. In all cases, $P_{EM}(\mathbf{x})$ is inconsistent with $P_{T_G}(\mathbf{x})$ when the sampling error is estimated using the leave one out procedure (method I). In the case of bootstrapping (method II), whether or not the confidence intervals overlap depend on the region and the temporal average of the pattern. For instance for Southern South America in the summer (4) the non linearity is only marginally larger than the internal variability uncertainty for $T_G = 1.5°$.

Table 1 lists the difference between the two patterns at each spatial location $\mathbf{x}$, $(P_{T_G}(\mathbf{x}) - P_{EM}(\mathbf{x}))$, measured in terms of the standard deviation of the ensemble of patterns $P_{EM}(\mathbf{x})$ for methods I and II, and when $(P_{T_G}(\mathbf{x})$ is calculated as in Case I. For method I, the difference between $P_{T_G}(\mathbf{x})$ and $P_{EM}(\mathbf{x})$ ranges between 3 and more than 40 standard deviations of the ensemble of patterns $P_{EM}(\mathbf{x})$. When using bootstrapping the differences between $P_{T_G}(\mathbf{x}$ and $P_{EM}(\mathbf{x}))$ are reduced, but for instance for Southern Europe in the summer still range between 1 and 5 standard deviations.

Figures 5, 6 and 7 are similar to Figures 2, 3 and 4 respectively, but when computing the patterns P_{T_G} as in Case II described above, i.e., using all the temperature change fields $T_m(\mathbf{x})$ for the models m whose global mean temperature T_m^G falls within the interval $T_G \pm 0.01$. As expected, the values of $P_{T_G}(\mathbf{x})$ depend on the interval chosen. The size of the ensembles around each temperature threshold T_G is smaller than in Case I ($21, 11$, and 18 versus $59, 38$, and 40 for $T_G = 1.5°, 2.0°$ and $2.5°$ respectively), increasing the sampling uncertainty as illustrated by the larger confidence intervals in this case.

Table 2 shows that the percentage difference between $P_{T_G}(\mathbf{x})$ and $P_{EM}(\mathbf{x})$ not only varies with region, season and T_G, but also with the choice of interval around T_G. While the nonlinearity is smaller than about 10% for Southern Europe for all time intervals and values of T_G, it can be much larger for Central North America where it reaches 27% in Case II in the winter. Whether

or not one can consider that the linearity assumption is violated in this case will presumably depend on these margin of errors being tolerable (or not) for the particular application.Presumably at smaller spatial scales, differences between $P_{T_G}(\mathbf{x})$ and $P_{EM}(\mathbf{x})$ will be larger due to non the linearity of the response at local scales, however confidence intervals might also increase, making it difficult to estimate how the results will be in those more impacts relevant cases.

2 Sensitivity of the results to the choice of baseline model and region

This section presents results for the risk of overshooting a temperature change threshold for different regions and choice of baseline model. The goal is to show that the inconsistency between the estimated risk of overshooting the threshold obtained from the MR and the PSR ensembles is not an artifact of the choice of baseline model M, nor particular to a region.

Figure 8 is analogous to figure 1 of the paper but for Central North America. The top and middle panels show the I.C. (MR) and the pattern scaled (PSR) ensembles' projections for boreal summer temperature change as a function of time. The MR projections are continuous time series for the period 1900-2079, but are plotted here as overlapping thirty years periods to facilitate the comparison with the PSR ensemble. The baseline anomalies of model M are shown in the first thirty year period for reference (green line). It is clear that, by construction, the year-to-year variability of the PSR ensemble follows the variability of model M anomalies, independently of the time slice considered (i.e., independently of s). This explains why the actual model M projections (blue line in middle panel) are not always contained within the range of the PSR ensemble.

The bottom panel of the figure shows the the risk of a heat wave occurrence, estimated as the fraction of model runs that overcome a given threshold (3.0° for Central North America for instance). For clarity, only decadal and thirty year means are shown in the figure. In both cases the results from the PSR ensemble do not coincide in general with the MR ensemble. For decadal averages , the results also vary depending on which time slice is used to calculate the pattern. For instance the risk of overshooting the threshold in the 2020-2029s is estimated to be 44%, 27% or 33% for Central North America by the PSR ensemble, depending on whether the spatial pattern P_s is calculated using the thirty years periods 2000-2029,2010-2039 or 2020-2049. The true risk estimated by the MR ensemble is 33% .

How significant are the differences between the true and the pattern scaling estimated risks in the context of sampling variability is not easy to estimate in general. However, in the context of the perfect model scenario, given that the IC ensemble has 32 members, any error smaller than about 3% is not significant, since the smallest possible frequency is $1/32 = 0.03$. In this sense,

for thirty year means the two ensembles (MR and PSR) are indistinguishable, with errors smaller than 1% for Central North America.

Figures 9 and 10 illustrate the results for the same two regions but a different random choice of the model used as baseline (model M). As expected, the two ensembles are not consistent with each other either, showing that this is not an artifact of the choice of baseline.

Note in passing that the inconsistencies signaled above will occur independently of whether or not the linear assumption evaluated in section 1 is correct. If the linear assumption is violated, then pattern scaling should not be used. But even if the linearity assumption was valid, the method described above produces inconsistent results for inter annual and inter decadal variability for the regions analised.

3 Sensitivity of the results to the sampling of inter annual variability

As discussed in section 3 of the paper, to obtain realizations of climate change superimposed to year-to-year variability, the spatial pattern of change is added to the observed time series of anomalies from climatology over the baseline period $T_{\text{baseline}}(\mathbf{x}, im, y)$, to yield the pattern scaled projections

$$T^*(\mathbf{x}, im, s, y) = P_s(\mathbf{x}, im)\ \tilde{T}(y) + T_{\text{baseline}}(\mathbf{x}, im, y) \tag{4}$$

where s denotes the time slice, y runs over the years within that time slice, and im labels month or seasons. In the previous section, T_{baseline} was represented by the baseline time series of a randomly chosen model M. Here, the effect of considering random permutations of this time series is explored; the underlying idea being that a better sampling of the internal variability can be achieved by randomly sampling segments of the baseline time series of anomalies.

Figures 11 and 12 show the results when employing the same baseline model as in figures 1 in the paper and 8 in this supplement, but constructing the PSR ensemble using sixty randomly permuted versions of the baseline time series. The permuted time series are obtained by randomly permuting each month separately.

The figures show that the pattern scaled ensembles increase their ranges, and for both regions the time series of the model chosen for the baseline is included within the range of the ensemble. For decadal averages, the results also vary depending on which time slice is used to calculate the pattern, but the differences are smaller than previously, becoming indistinguishable from the minimum possible frequency difference, 1/32.

This is to be expected; the large discrepancies between decades in the examples discussed in the previous section and in the paper, were due to the fact that the same fixed baseline was added to the thirty year changes computed as $P_s(\mathbf{x}, im)\ \tilde{T}(y)$. When the baseline time series is randomly permuted, this problem is minimized.

For instance, as shown in figure 11 for Southern Europe, the risk of overshooting the threshold in the 2020-2029s is estimated to be 37%, 38% or 41% by the PSR ensemble, and the result depends on whether the spatial pattern P_s is calculated using the thirty years periods 2000-2029, 2010-2039 or 2020-2049. The true risk estimated by the MR ensemble is 39% . In the case of Central North America, figure 12, the PSR estimates are closer together (33%, 30% or 34%) and to the true value 33%. Errors for thirty year means remain comparable to the ones obtained when taking a fixed baseline. They range between -4% and 5% for both Southern Europe and Central North America. This is consistent with the fact that exploring the variability within each time slice can not change potential errors in the averages over the entire time slices.

Table 1 Estimation of errors induced by the linearity assumption: ratio between $(P_{T_G}(\mathbf{x}) - P_{EM}(\mathbf{x}))$ and the standard deviation of the $P_{EM}(\mathbf{x})$ ensemble obtained using methods I and II, for annual, and boreal winter and summer patterns. For each of the regions shown in figures 2 to 4 the rows from top to bottom correspond to $T_G = 1.5°, 2.0°$ and $2.5°$, and the intervals around these temperatures are chosen as in case I.

	Method I			Method II		
Region	annual	DJF	JJA	annual	DJF	JJA
Central	20.4	23.6	12.0	3.8	4.4	2.2
North	3.6	14.2	-14.2	0.7	2.6	-2.6
America	-4.0	-15.6	-2.9	-0.8	-2.9	-0.5
Southern	26.0	42.6	17.2	4.8	7.9	3.0
South	10.9	-20.4	22.0	2.0	-3.8	4.1
America	2.8	-20.2	30.4	0.5	-3.8	5.5
Southern	-8.6	-11.4	10.1	-1.6	-2.1	1.9
Europe	6.4	23.8	28.8	1.2	4.4	5.3
	6.7	14.1	6.3	1.2	2.6	1.2

Table 2 Estimation of errors induced by the linearity assumption: percentage difference $(P_{T_G}(\mathbf{x}) - P_{EM}(\mathbf{x}))/P_{EM}(\mathbf{x})$ when defining $P_{T_G}(\mathbf{x})$ as in case I (first three columns) and case II (last three columns) , and using method II to estimate the confidence intervals. For each of the regions, the rows from top to bottom correspond to $T_G = 1.5°, 2.0°$ and $2.5°$.

	case I			case II		
Region	annual	DJF	JJA	annual	DJF	JJA
Central	7.6	10.3	3.1	11.1	27.0	0.3
North	1.6	5.5	-3.8	14.3	23.2	6.4
America	-2.1	- 6.7	-0.8	-4.3	-14.7	0
Southern	4.5	5.7	3.8	11.8	10.5	16.1
South	2.2	-2.7	4.9	2.0	1.0	-5.2
America	0.9	-2.7	6.6	3.9	-2.0	11.3
Southern	-2.4	-2.6	1.4	0.3	-2.5	3.6
Europe	1.4	6.3	4.4	0	-3.5	7.7
	1.5	3.5	1.0	-3.8	-7.7	0

References

1. Efron, B.: The jacknife, the bootstrap, and other resampling plans. Society of Industrial and Applied Mathematics, CBMS-NSF Monographs 38 (1982)
2. Ruosteenoja, K., Tuomenvirta, H., Jylh, K.: Gcm-based regional temperature and precipitation change estimates for europe under four sres scenarios applying a super-ensemble pattern-scaling method. Climatic Change **81**, 193–208 (2007)

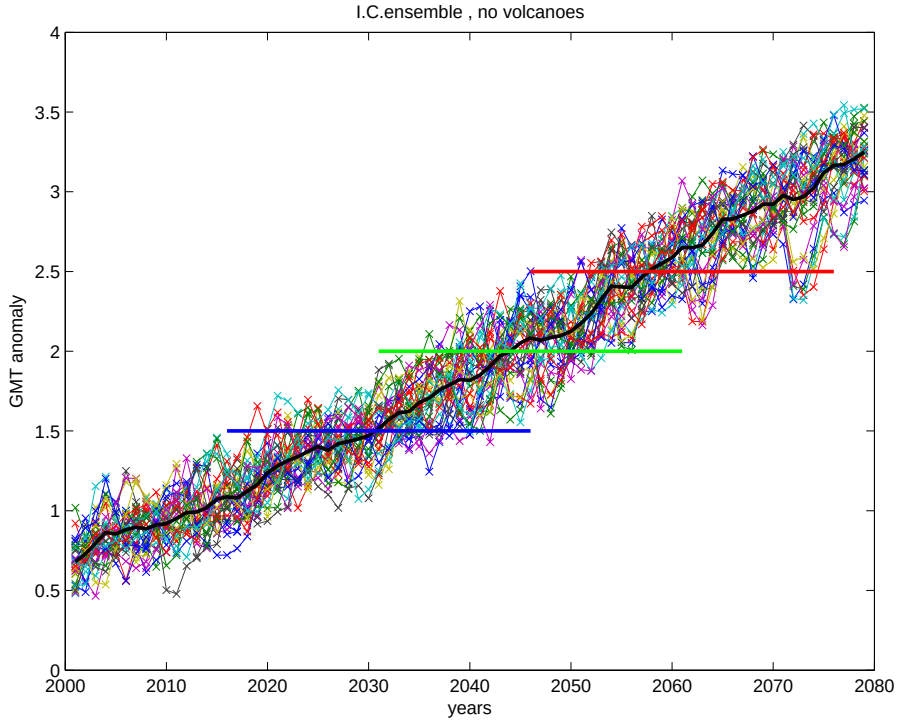

Fig. 1 Time series of GMT anomalies for the 32 member I.C. ensemble (colored thin lines). The thick black line is the ensemble mean. The blue, green and red thick horizontal lines indicate $T_G = 1.5°, 2.0°$ and $2.5°$. The linearity assumption is verified if the pattern obtained using the ensemble mean times series, $P_{EM}(\mathbf{x})$, is identical to the patterns obtained using the models' temperature fields across the horizontal thick colored lines, $P_{T_G}(\mathbf{x})$

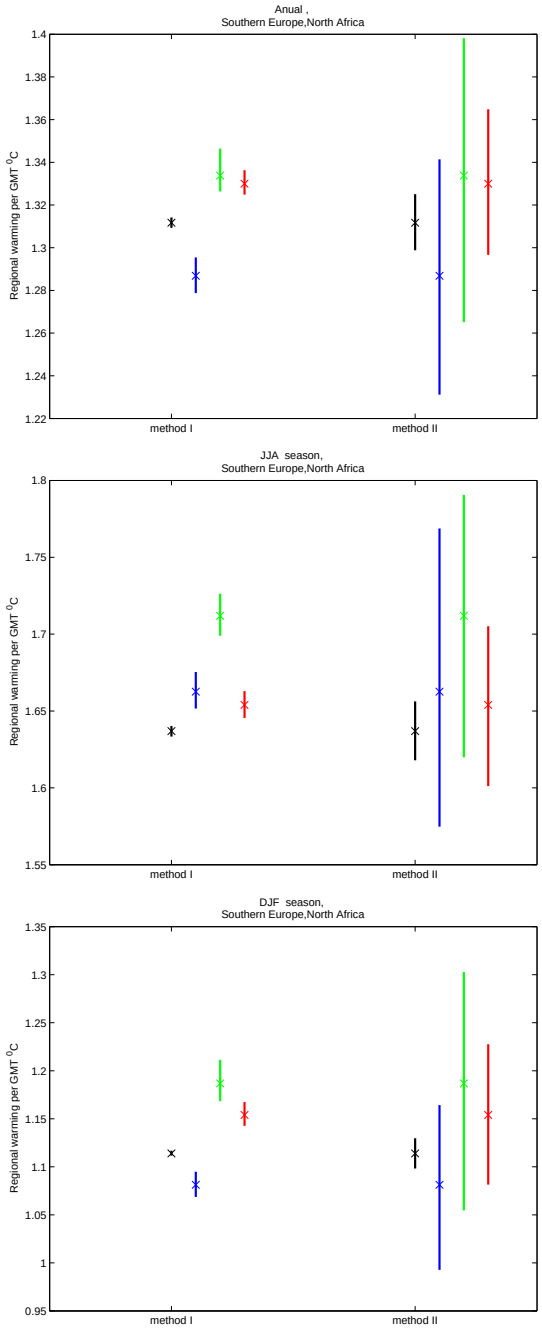

Fig. 2 Patterns P_{EM} and P_{T_G} for Southern Europe. Black symbols correspond to P_{EM} and colored symbols correspond to three different values of P_{T_G} for $T_G = 1.5°$ (blue), $2°$ (green) and $2.5°$ (red). Method I and method II correspond to two different ways of estimating the error bars (see text), 90% confidence intervals are plotted. The top, middle and bottom panels show annual, JJA and DJF patterns respectively.

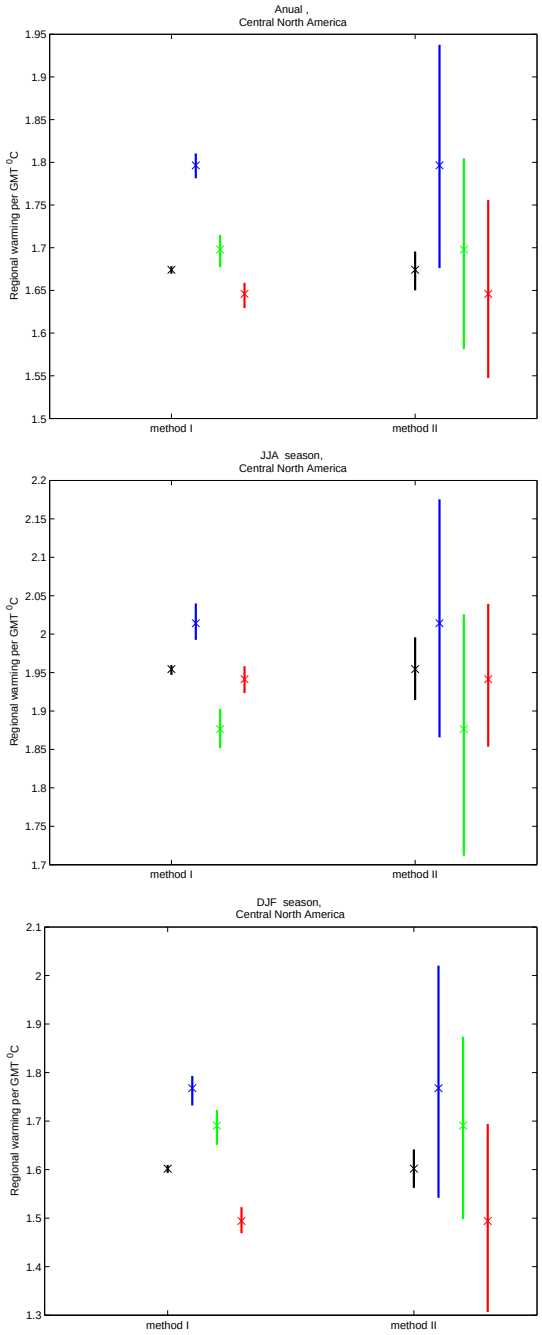

Fig. 3 Same as figure 2 but for Central North America.

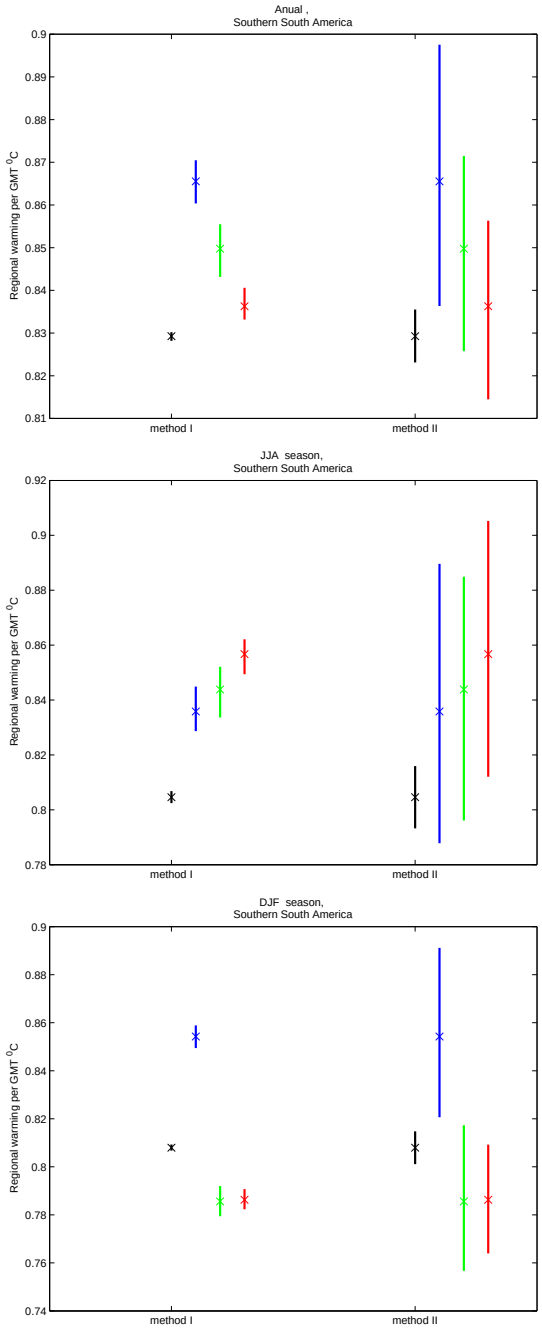

Fig. 4 Same as figure 2 but for Southern South America.

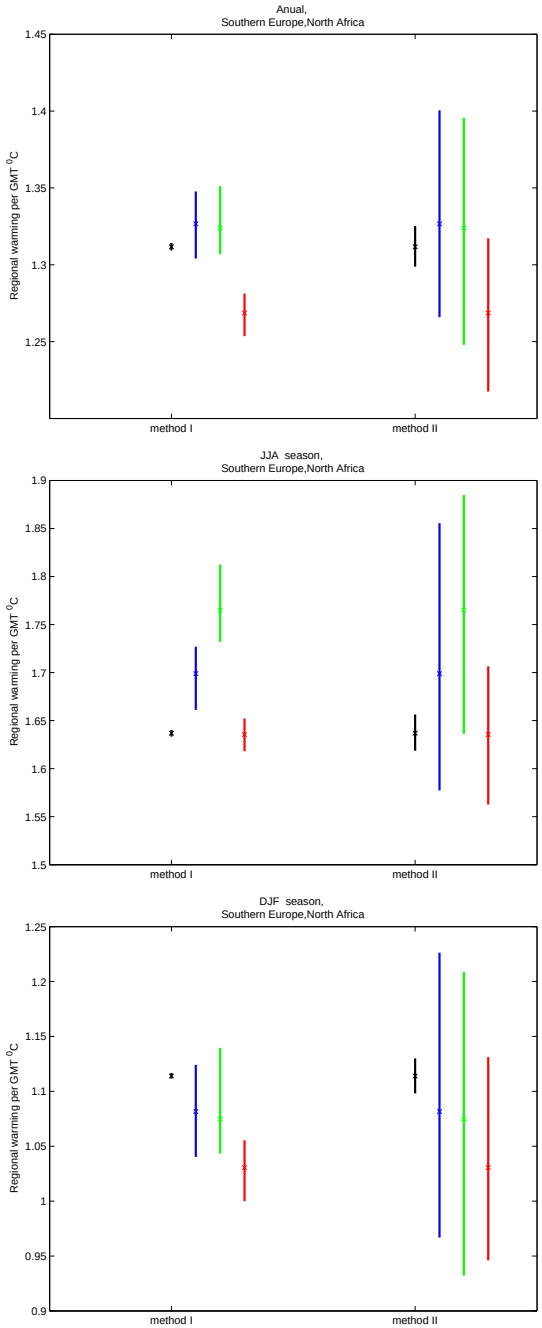

Fig. 5 Southern Europe. Same as figure 2 but for P_{T_G} patterns computed using fixed threshold $T_G \pm 0.01°$.

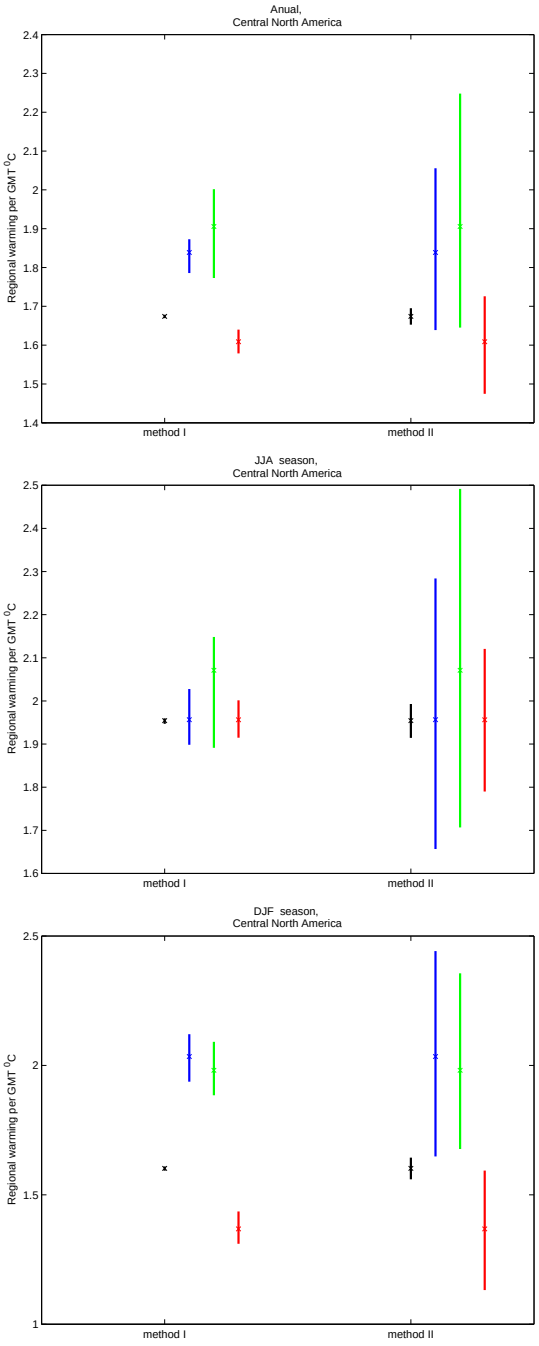

Fig. 6 Central North America. Same as figure 3 but for P_{T_G} patterns computed using fixed threshold $T_G \pm 0.01°$.

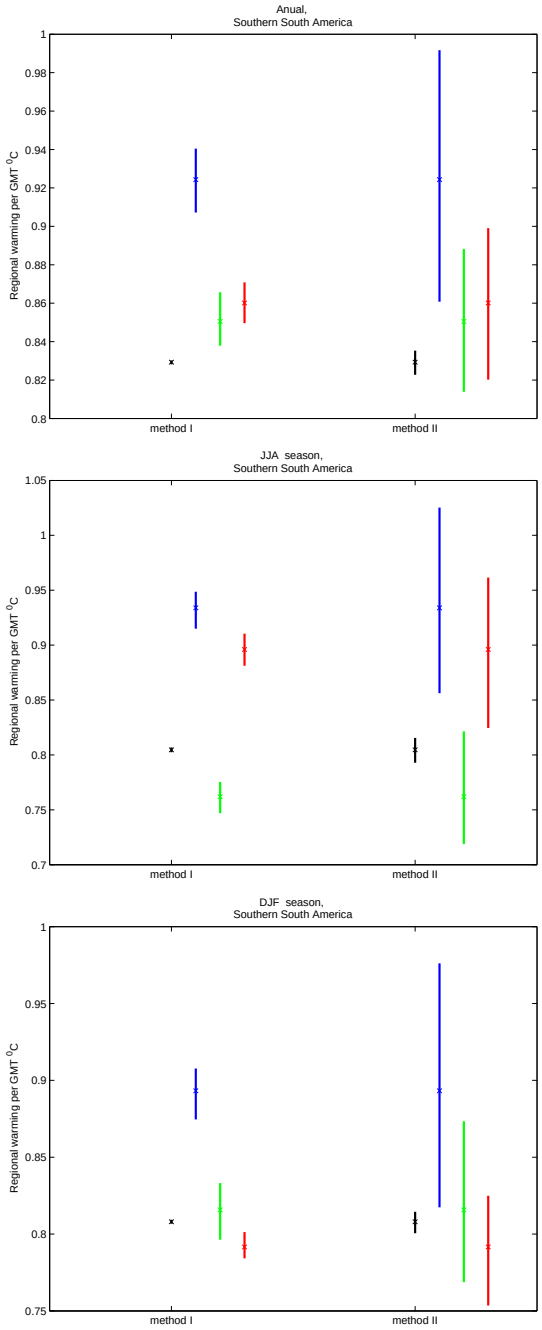

Fig. 7 Southern South America. Same as figure 4 but for P_{T_G} patterns computed using fixed threshold $T_G \pm 0.01°$.

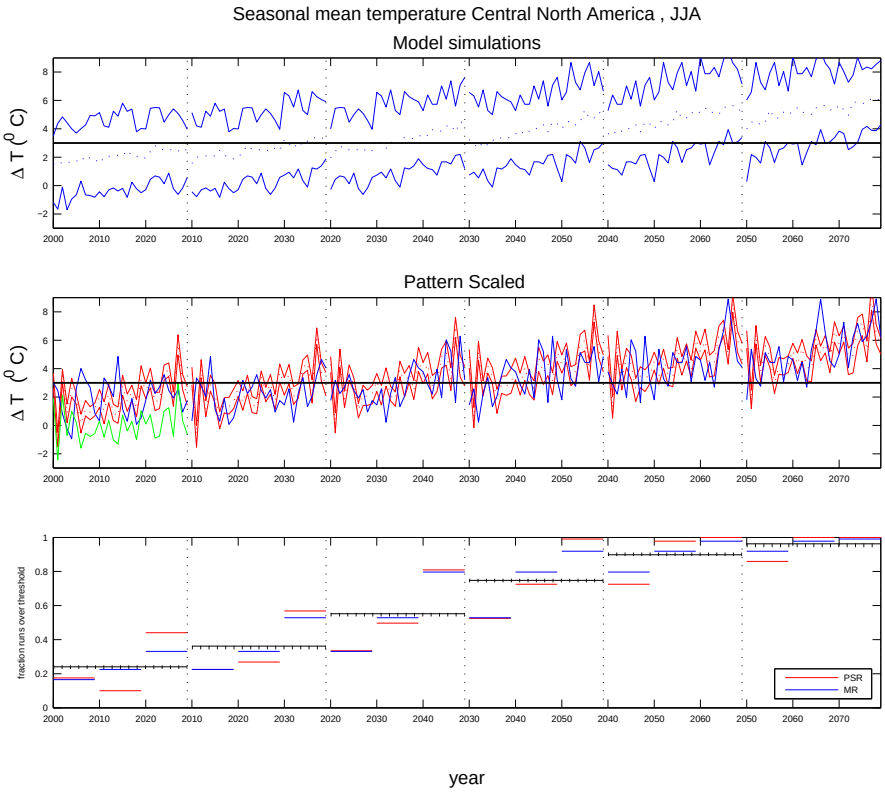

Fig. 8 Central North America projections for summer warming and risks of heat waves. **Top and middle panels**: The I.C. (top) and PSR (middle) ensembles' projections for summer temperature change are plotted as a function of time. The thick horizontal line indicates the 2.3° threshold. In these two panels the two solid lines correspond to the ensemble range and the dotted line is the 50% percentile. Notice that for the MR the projections are continuous time series for the period 1900-2079, but are plotted here as overlapping thirty years periods to facilitate the comparison with the PSR ensemble. The blue line in the middle panel correspond to the I.C. model run randomly sampled to be used as surrogate for observed anomalies (model M), and the green line (shown for reference in the first thirty years period) corresponds to the baseline (1961-1990) anomalies of model M used to construct the PR ensemble. Notice that the pattern scaled ensemble does not completely enclose the trajectory of model M (blue line) as expected by construction. **Bottom panel**: the risk of a heat wave occurrence, estimated as the fraction of model runs that overcome the 2.3° threshold, as quantified by the MR (blue) and the PSR (red) ensembles. Solid colored lines indicate fractions of runs over threshold for decadal means, and black solid (MR ensemble) and black dashed (PSR) lines correspond to thirty years means. Notice that the MR projections of changing risk are continuous time series for the period 1900-2079, but are plotted here as overlapping thirty years periods to facilitate the comparison with the PSR ensemble

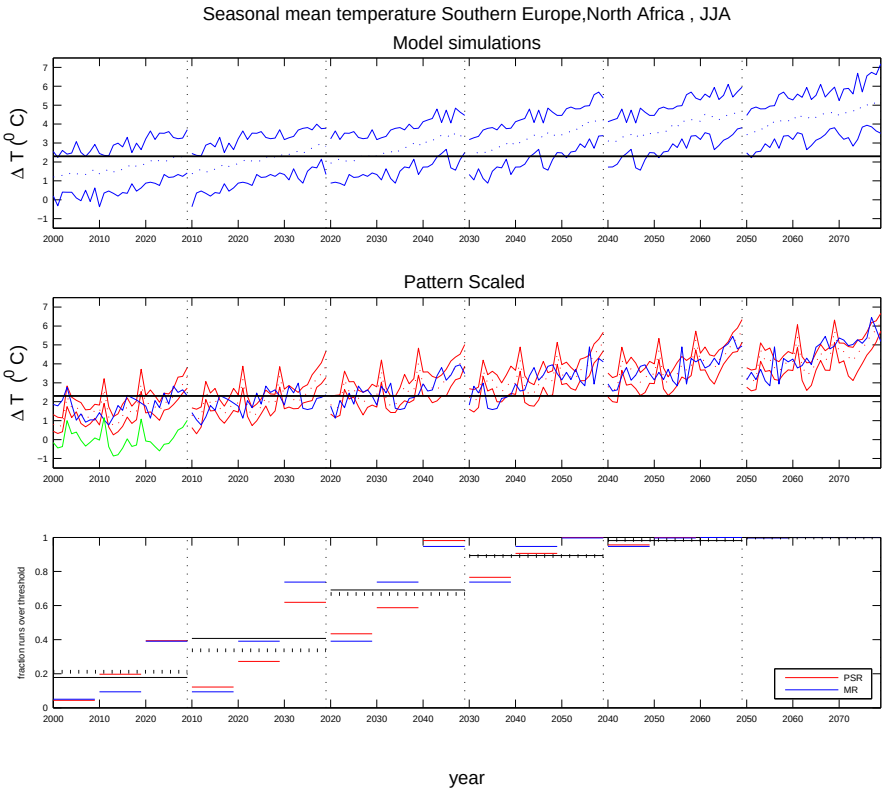

Fig. 9 Same as figure 8 for Southern Europe but a different model M

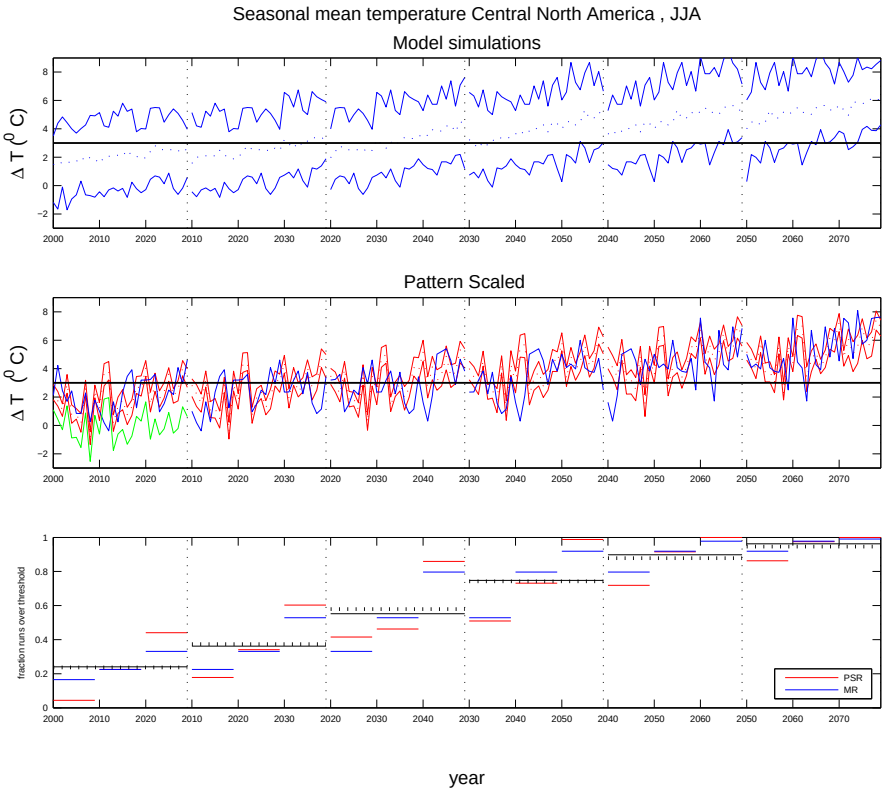

Fig. 10 Same as figure 8 for Central North America, but a different model M

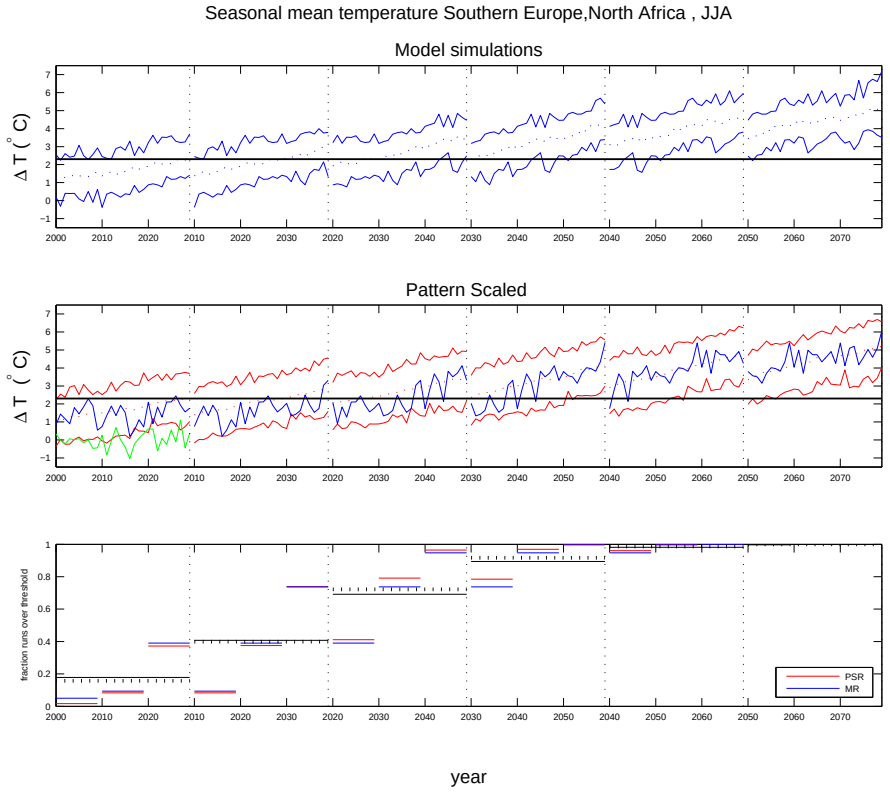

Fig. 11 Southern Europe projections for summer warming and risks of heat waves. The pattern scaled ensemble in the middle pannel is obtained by adding the spatial patterns of change $P_s(\mathbf{x}, im)\ \tilde{T}(y)$ to randomly permuted versions of the baseline time series $T_{\text{baseline}}(\mathbf{x}, im, y)$ from the same model M as in figure 1 in the paper.

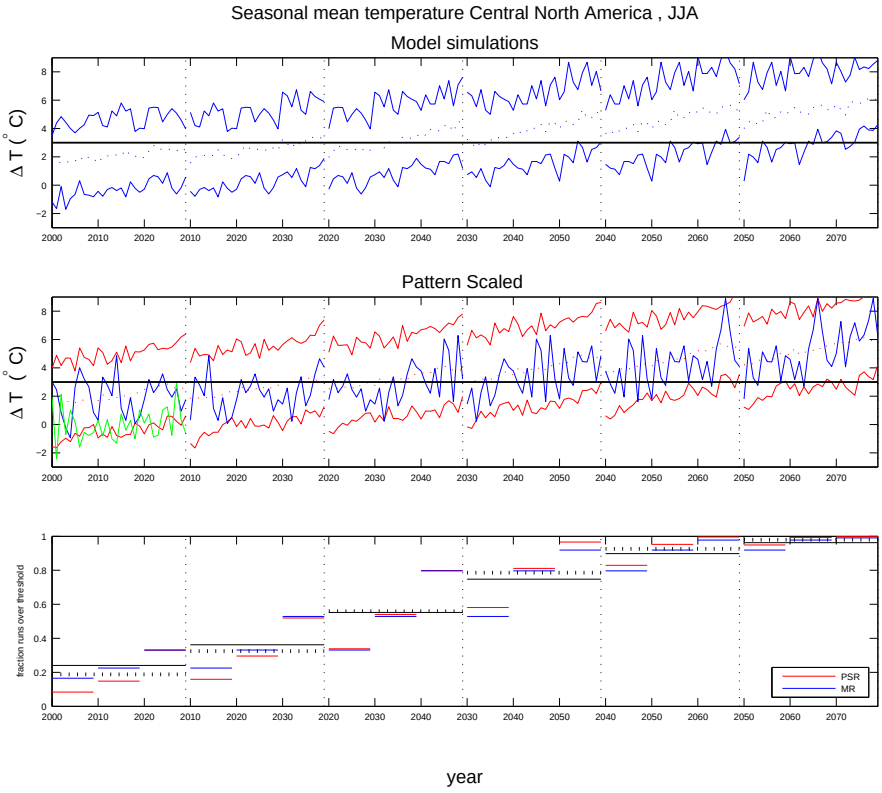

Fig. 12 Central North America projections for summer warming and risks of heat waves. The pattern scaled ensemble in the middle pannel is obtained by adding the spatial patterns of change $P_s(\mathbf{x}, im)\ \tilde{T}(y)$ to randomly permuted versions of the baseline time series $T_{\text{baseline}}(\mathbf{x}, im, y)$ from the same model M as in figure 8.