INTERNATIONAL JOURNAL OF CLIMATOLOGY
*Int. J. Climatol.* (2013)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.3659

RMetS
Royal Meteorological Society

# The potential of pattern scaling for projecting temperature-related extreme indices

A. Lustenberger,* R. Knutti and E. M. Fischer
*Institute for Atmospheric and Climate Science, ETH, Zurich, Switzerland*

**ABSTRACT:** Pattern scaling can be used to linearly relate changes in extreme indices to changes in the annual or seasonal mean temperature. This study demonstrates the skills and limitations of two often used pattern scaling approaches in filling-in gaps in the time series of six temperature-related extreme indices. The extreme indices over Europe are derived from daily temperature output of 12 regional climate models of the multi-model project ENSEMBLES. The response pattern is estimated using one of the two future time periods (2021–2050 or 2070–2099) and the reference period (1961–1990). The simulated values from the remaining future time period are used for evaluating the skills. Both pattern scaling approaches perform reasonably well particularly for percentile-based and over most of the regions also for fixed temperature indices. Uncertainties due to internal variability can be large if the time period used for estimating the response pattern is close to the reference period. Limitations of pattern scaling due to violations of the linearity assumption are related to the shape of the temperature distribution. As a result, differences in the skills among the extreme indices can be related to the magnitude and shift direction of the whole temperature distribution. Therefore, skills for estimated extreme indices derived from the upper tail of the underlying temperature distribution are generally high. Over some areas, linear regression models used in this study are not appropriate statistical models because of the bounded and discrete nature of the data. Alternative pattern scaling methods such as, for instance, the logistic regression model leads to improvements over particular areas but not over the whole integration area. 

KEY WORDS pattern scaling; temperature extremes; regional climate model; ENSEMBLES; missing values

*Received 13 March 2012; Revised 17 December 2012; Accepted 17 December 2012*

## 1. Introduction

Changes in the frequency or intensity of extreme events as a result of a changing climate would have major impacts on society and the environment. In this context, not only climate scientists but also different impact communities and end-users such as, for instance, the reinsurance companies are highly interested in trends in the frequency, intensity and duration of extreme events and the understanding of the driving mechanisms. From a statistical point of view, their long return periods are a major challenge in the assessment of extreme events. Consequently, there are in general not enough observed or simulated extreme events to reliably estimate their statistical properties. Additionally, climate models still have difficulty simulating the typically high nonlinear actions across scales and small-scale nature of extreme events. To avoid some of these problems we here use a set of extreme indices with return periods in the order of weeks. Such moderate definitions of extremes include events that do not correspond to the extremeness of events that are generally perceived as weather extremes often associated with high impacts. However, the moderate definitions used here and thus the short return periods ensure a sufficiently large sample size for robust trend analysis.

According to the extreme value theory, these events are statistically not extreme because they do not follow the extreme value distributions. Extreme indices are used for assessing and monitoring changes in extremes on global scale (Alexander *et al*., 2006) or in regional trend studies (Moberg *et al*., 2006; Toreti and Desiato, 2008). We here use the definitions from the European Climate Assessment and Dataset (ECA&D) initiative (http://eca.knmi.nl/indicesextremes/index.php).

Simulations performed with regional climate model (RCM) or global climate model (GCM) are an important tool in climate science to assess and understand changes in the climate system and their driving processes. However, the associated costs limit the availability of simulated data for some emission scenarios or time periods. This is not only an issue for climate scientists but also for the impact community using these simulations as input for their impact models. A widely used approach to fill in these gaps is pattern scaling, a method that provides climate change scenarios for time periods or emission scenarios for which no simulations are available.

* Correspondence to: A. Lustenberger, Institute for Atmospheric and Climate Science, ETH, Universitätstrasse 16, 8092 Zurich, Switzerland.
E-mail: andreas.lustenberger@env.ethz.ch

The underlying idea is that changes in a regional variable are linearly related to large-scale or global mean temperature changes. A forcing and time-independent response pattern determines the variable of interest from the selected set of predictors. A fast and easy way to estimate the response pattern is the time-slice method (Mitchell, 2003). The response pattern is then given by the differences in the estimated means of the variable of interest scaled by the mean temperature change between two distinct time periods or emission scenarios. However, any deviation from the underlying Gaussian assumption or, more general, from symmetry concerning the underlying probability density functions (PDF) leads to biases in the response pattern. A more robust method is to fit a linear regression using the least-squares method because this approach takes all available data points into account, minimizing the mean-squared error (Mitchell, 2003). Mitchell and Hulme (1999) and Huntingford and Cox (2000) applied the least-squares method to a sequence of decadal means. Ruosteenoja *et al*. (2007) used the least-squares approach in order to fit several simulations of four different emission scenarios for a specific GCM on the global mean temperature change simulated by a simple climate model. Kennett and Buonomo (2006) scaled the time-varying mean and standard deviation of the RCM with the smoothed global mean surface air temperature from the driving GCM. In the framework of the ENSEMBLES project, Kendon *et al*. (2010) described how pattern scaling can be used to increase the number of GCM–RCM chains.

The main idea of regression is to estimate the conditional mean of a variable of interest. This implies that a trend-induced change in the location parameter of the dependent variable has to be related to the trend-induced change in the mean of the explanatory variable (Simolo *et al*., 2010). But especially extremes and extreme indices are further sensitive to changes in the scale or shape parameter of the underlying PDF (Katz and Brown, 1992). In this context, the objective of this study is to evaluate the applicability of the time-slice and least-squares method to a set of temperature-related extreme indices using the IPCC SRES A1B emission scenario which runs from a subset of GCM–RCM pairs in the framework of the ENSEMBLES project (Van Der Linden and Mitchell, 2009). The pattern scaling methods are used in this study to fill in gaps in the time series of extreme indices. As there is mainly one emission scenario available in the ENSEMBLES project, the skills of the scaling methods are therefore tested using different time periods rather than different emission scenarios. The examination of the underlying assumptions such as, for instance, the linearity assumption is crucial.

The structure of this study is as follows: Section 2 introduces the RCMs and the set of temperature-related extreme indices. In Section 3, the least-squares and time-slice method and their associated assumptions are discussed. The results are presented in Section 4 and discussed in Section 5. The resulting conclusions are the topic of Section 6.

## 2. Data

### 2.1. Extreme indices

In this study, six temperature-related extreme indices are used, each of them derived from daily 2-m temperature data. The extreme indices are defined as follows:

- Number of frost days (FD): If TN is the daily minimum temperature, FD is the number of days per year when $TN < 0\ ^{\circ}C$.
- Number of summer days (SU): SU counts the number of days per year when the daily maximum temperature $TX > 25\ ^{\circ}C$.
- Percentage of warm days (TX90P) and cold days (TX10P) per year: For each calendar day of the reference period 1961–1990 the 90th and 10th percentile of the daily maximum temperature is computed using a five day time window centred at the calendar day.
- Percentage of warm nights (TN90P) and cold nights (TN10P) per year: Same as above but for daily minimum temperature.
- Cold-spell days index (CWFI): If TG is the daily mean temperature and TG10P the calendar day 10th percentile of a five day window centred at the calendar day, CWFI counts the number of days per year when TG < TG10P for at least six consecutive days.
- Warm-spell days index (HWFI): If TG is the daily mean temperature and TG90P the calendar 90th percentile of a five day window centred at the calendar day, HWFI counts the number of days per year when TG > TG90P for at least six consecutive days.

All six extreme indices are computed with the Climate Data Operators (CDO). CDO is an open source tool developed by the Max-Planck-Institute (https://code.zmaw.de/projects/cdo/) and provides a set of operators for analysing climate model outputs. The extreme indices implemented in CDO are European Climate Assessment & Dataset (ECA&D) indices. However, six of the extreme indices (FD, SU, TX10P, TX90P, TN10P and TN90P) used in this study follow the recommended definitions by the Expert Team on Climate Change Detection and Indices (ETCCDI). HWFI and CWFI have definitions which are very similar to the warm-spell duration index (WSDI) and cold-spell duration index (CSDI) suggested by ETCCDI. The main difference is that the daily mean temperature is used to compute CWFI and HWFI. In contrast to this, WSDI and CSDI employ daily maximum and minimum temperature, respectively. The suggested extreme indices are day-count indices based on fixed (FD, SU) or percentile (TX90P, TX10P, TN90P, TN10P, HWFI and CWFI) thresholds. The fixed extreme indices make a spatial comparison difficult because they do not sample the same part of the underlying probability distribution of temperature at different sites. The motivation of using extreme indices based on fixed thresholds is that they

 

Table I. The ENSEMBLES regional climate models (RCM) with the same rotated pole grid and available simulations for the time period from 1951 to 2099.

| Institution | RCM model | Driving GCM |
|---|---|---|
| Irish National Meteorological Service (C4I) | RCA3 | HadCM3Q16 |
| Danish Meteorological Institute (DMI) | DMI-HIRHAM | ECHAM5 |
| Danish Meteorological Institute (DMI) | DMI-HIRHAM | ARPEGE |
| Swiss Federal Institute of Technology (ETH) | CCLM | HadCM3Q0 |
| Royal Netherlands Meteorological Institute (KNMI) | RACMO | ECHAM5 |
| Met Office Hadley Centre (HC) | HadRM3Q0 | HadCM3Q0 |
| Met Office Hadley Centre (HC) | HadRM3Q3 | HadCM3Q3 (low sensitivity) |
| Met Office Hadley Centre (HC) | HadRM3Q16 | HadCM3Q16 (high sensitivity) |
| Swedish Meteorological and Hydrological Institute (SMHI) | RCA | ECHAM5 |
| Swedish Meteorological and Hydrological Institute (SMHI) | RCA | BCM |
| Swedish Meteorological and Hydrological Institute (SMHI) | RCA | HadCM3Q3 |
| Max-Planck-Institute for Meteorology (MPI) | REMO | ECHAM5 |

are often associated with observed impacts. For spatial comparison the percentile-based extreme indices are more suitable because they have a constant rarity in context of the local climate. SU and FD do not take the annual cycle into account in their definitions. This means that in regions with a pronounced annual cycle one can expect that the seasons affect the internal variability of these two indices.

### 2.2. ENSEMBLES regional climate models

The daily maximum, minimum and mean 2-m temperatures are simulated by different one-way nested GCM–RCM chains. The GCM–RCM pairs are defined in the framework of the ENSEMBLES project (Van Der Linden and Mitchell, 2009). Due to limited computational resources, the ENSEMBLES community decided to use mainly the SRES A1B emission scenario. In this study, a set of 12 GCM–RCM chains is used with a common grid and transient runs covering the time period from 1951 to 2099. The horizontal resolution of the RCMs is about 25 km. Only grid points exhibiting a land area fraction of more than 50% are considered in this study. Table I gives an overview on the 12 GCM–RCM pairs, i.e. the institutions, the name of the RCM and the name of the driving GCM.

## 3. Methods

### 3.1. Estimation of the response pattern

There are two ways to estimate the response pattern, the time-slice approach (Mitchell, 2003) and the least-squares approach (Mitchell, 2003; Huntingford and Cox, 2000). The annual mean of the daily mean, daily maximum and daily minimum 2-m temperatures are used as explanatory variables. Additionally, in case of the FD and SU index, the seasonal mean summer (JJA) and mean winter (DJF) temperatures are used. In cases where the seasonal mean temperature is employed, the extreme indices are computed per season. The anomalies for each variable are computed by subtracting the mean of the reference period (1961–1990) from the mid-century (2021–2050) and late-century (2070–2099) period mean. For the least-squares approach, a linear regression model is used with an unknown regression coefficient $\beta_1$ and the intercept $\beta_0$ which is assumed to be zero. The errors $\varepsilon_i$ have to satisfy the Gauss–Markov conditions (Plackett, 1950). The basic idea of the least-squares method is to minimize the residual sum of squares with respect to the regression coefficient $\beta_1$. Therefore, $\beta_1$ can be estimated in the following way:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})\,(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \tag{1}$$

The subscript $i$ denotes all years of the reference and the scenario period. For each grid point the significance of the assumed linear trend is tested with a simple $t$-test of the coefficient $\beta_1$ with the significance level $\alpha = 5\%$. If the Gauss–Markov conditions are not violated then $T$ is distributed like Student's $t$ with n−2 degrees of freedom. The advantage of the least-squares approach relative to the time-slice approach is that all available data are used. Additionally, due to minimizing a quantity measuring the errors we can be sure that, if all assumptions are fulfilled, this approach fits the data better than the time-slice method.

Another possibility to estimate the slope coefficient $\beta_1$ is the time-slice method. This approach estimates $\beta_1$ in the following way:

$$\widehat{\beta}_1 = \frac{\overline{y}_{\text{Scenario}} - \overline{y}_{\text{Reference}}}{\overline{x}_{Scenario} - \overline{x}_{\text{Reference}}} \tag{2}$$

$\beta_1$ is only estimated when there is a significant trend-induced change of the mean at the 5% significance level. In this context, a simple $t$-test is applied with the null hypothesis $H_0$ that the means are equal.

The scaling factor $\beta_1$ is estimated for each grid point and for each of the 12 GCM–RCM pairs using both pattern scaling approaches. In addition, $\beta_1$ is only estimated for a particular grid point when the associated significance test rejects the null hypothesis.

 

### 3.2. Evaluation of the estimation skills

The response pattern is derived using one of the two future time periods and the reference period. The estimated response pattern is (1) used to interpolate the extreme indices for the mid-century period 2021–2050 based on information from the control period and the late-century 2070–2099, or to (2) extrapolate them for the late-century 2070–2099 based on information from the control period and the mid-century, respectively. The simulated values from the remaining time period are used for evaluating the skill of the extreme index estimates. The motivation of this study is to stay on the level of the RCMs, i.e. we assume that we have only information from the 12 GCM–RCM chains presented in Section 2.2. An alternative procedure would be to use the global mean temperature change as predictor or regional information derived from statistical downscaling. However, because the information of the RCM are more closely related to regional climate simulated by the GCM–RCM chains, the choice of RCM-based information delivers better insight into the limitation of pattern scaling applied to extreme indices.

As a measure of skill, two different statistical quantities are used here. First, the root mean squared error (RMSE). It is important to keep in mind that the RMSE is sensitive to outliers. Another statistical quantity is the Pearson's correlation coefficient $\rho$. The correlation coefficient measures the degree of correspondence between the estimated and simulated values. The squared value $\rho^2$ expresses the fraction of the variance in the simulated values explained by the estimated linear regression. However, $\rho^2$ does not take into account any bias. Therefore, a high value of $\rho^2$ can still have a large systematic error. For this purpose, the RMSE will be used additionally in order to quantify potential biases. Similar as the RMSE, $\rho^2$ is sensitive to outliers.

## 4. Results

### 4.1. Skill of the scaled extreme indices

Figure 1 shows in the top panel the mean TX90P values simulated with the HadCM3Q16-HadRM3Q16 model and in the bottom panel the values fitted with the least-squares approach for the time slice 2070–2099 using the yearly mean daily maximum temperature as explanatory variable. Over central and northern Europe, there is an underestimation of the fitted mean TX90P values. In contrast to this, the TX90P values over southern Europe are in relatively good agreement with the simulated ones. The figure representing the fitted values implies some spatial dependency of the skills which will be discussed later. Figure 2 shows the $\rho^2$ and RMSE values for the estimated TX90P ((a), (b)) and TX10P ((c), (d)) indices over the integration area of each GCM–RCM pair. Changes in the yearly mean daily maximum temperature are used as explanatory variable. The range in $\rho^2$ and RMSE across grid points is depicted with the empirical 25 th and 75th percentile.

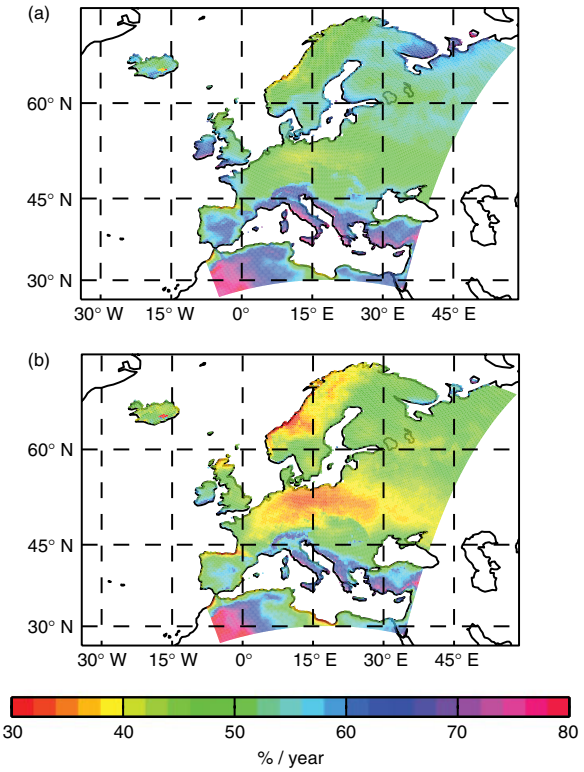


Figure 1. Mean TX90P for the time slice 2070–2099 (a) simulated with the HadCM3Q16-HadRM3Q16 model and (b) fitted with the least-squares approach using the annual mean daily maximum temperature as predictor.

Figure 2(a) and (c) shows the skill values for the time period 2021–2050 (interpolation) whereas Figure 2(b) and (d) those for 2070–2099 (extrapolation). For both extreme indices, the least-squares approach performs better than the time-slice approach. The multi-model median RMSE values in case of the least-squares approach are for both time periods of interest about 23–57% lower than for the time-slice approach. In addition, the interpolation for the time period 2021–2050 yields substantially better results than the extrapolation for 2070–2099. The multi-model interquartile range, as a measure for the inter-model differences, is up to 50% lower for the least-squares approach and the time period 2021–2050 relative to the time-slice approach. When estimating the extreme indices for the time period 2070–2099, the inter-model differences in the RMSE values become larger. Especially for the TX10P index, the RCMs driven with the high climate sensitivity GCM can be characterized as outliers in terms of the RMSE values.

Concerning the $\rho^2$ values, Figure 2 shows that for both pattern scaling approaches the correlation between the simulated and fitted TX90P values are between 1.5–3 times higher than for the TX10P index. Additionally, the multi-model interquartile range is in almost all cases smaller for the TX90P index. This behaviour implies that, given increasing mean temperatures, estimating extreme indices associated with the upper tail of the underlying temperature distribution (e.g. TX90P) seems to work generally better than for those in the lower tail. This

 

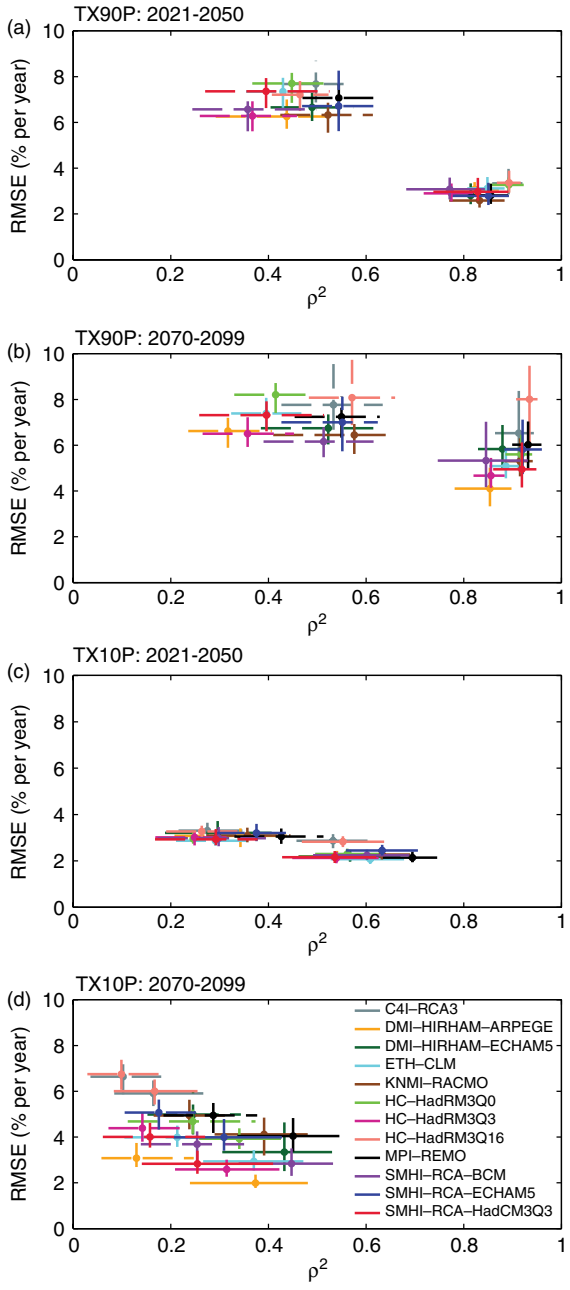


Figure 2. Skill of the pattern scaling approach for all 12 GCM–RCM chains expressed as squared correlation $\rho^2$ and RMSE values. Skills are shown for TX90P in (a) 2021–2050, and (b) 2070–2090, and for TX10P in (c) 2021–2050, and (d) 2070–2099. The indices are estimated with the time-slice (dashed line) and least-squares approach (solid line). The annual mean daily maximum temperature is used as explanatory variable. The range of values over the integration area is represented with the empirical 25th and 75th percentile.

behaviour is due to the fact that the TX10P index approaches 0% per year as a result of the warming. Therefore, the relationship between the extreme index and mean temperature tends to flatten and thus violate the assumption of linearity. The same behaviour is observed for the TN10P and TN90P index (not shown) derived from the daily minimum temperature distribution. However, there are very small differences in the skill values between the TN10P and TX10P extreme index and the TN90P and TX90P index, respectively. In case of the fixed threshold based extreme indices (FD and SU; not shown), the multi-model median RMSE values are still lower for the least-squares approach than for the time-slice approach for the interpolation and extrapolation. Seasonal mean temperature as predictor for FD and SU leads to multi-model median RMSE values up to 50% smaller than for the corresponding annual based values. For both pattern scaling approaches, this decrease is for the interpolated estimations larger than for extrapolated values. For the multi-model $\rho^2$ values, an improvement up to 28% is evident mainly for the least-squares approach applied on FD. In general, better skills are found over northern Europe for FD and over the Mediterranean in case of SU. This is due to the fact that over northern Europe trends in annual mean temperatures are dominated by trends in the seasonal DJF mean temperature while over southern Europe trends in the JJA mean temperature have a larger influence.

For yearly FD, the $\rho^2$ values (multi-model median and interquartile range) are comparable with the values of the TX10P and TN10P values. In case of SU, the multi-model median values of this particular skill measurement are lower and the multi-model interquartile range is in almost all cases higher than for the values of the TX90P and TN90P index. In addition, the higher interquartile range values for each GCM–RCM pair in case of FD and SU imply a higher spatial dependency of the $\rho^2$ values relative to the quantile-based extreme indices.

The lowest correlation values (not shown) are found between CWFI and the annual mean temperature. The median value over the integration area of each GCM–RCM chain is between −0.5 and −0.56 with a maximum interquartile range of 0.11. The small inter-model differences give a hint that for this particular extreme index the linearity assumption is violated. In contrast to the CWFI, the skill values for the HWFI indicate larger systematic errors in the estimations but better correspondence with multi-model median $\rho^2$ values in the order of 0.35–0.75 (not shown).

To investigate the uncertainty induced by internal variability on the estimations of the slope parameter $\beta_1$, a 30-year moving window is used together with the fixed reference period 1961–1990. The moving time window is shifted by 1 year at each step and moves through the time period from 1991 to 2099. Figure 3 shows the estimated $\beta_1$ values for the TX90P index based on the time-slice method (blue line) and the least-squares method (red line) for a single grid point over northern Scandinavia from the ECHAM5-RCA model. The TX90P index is chosen because this extreme index does not exhibit any significant violation of the linearity assumption over the whole integration area. The green lines represent the range of estimated $\beta_1$ values for that particular grid point for all 12 GCM–RCM chains. The blue and red curves indicate that the uncertainty in the slope parameter estimates is largest when the moving window is close to the reference period. This implies that the warming signal is partly masked by

 

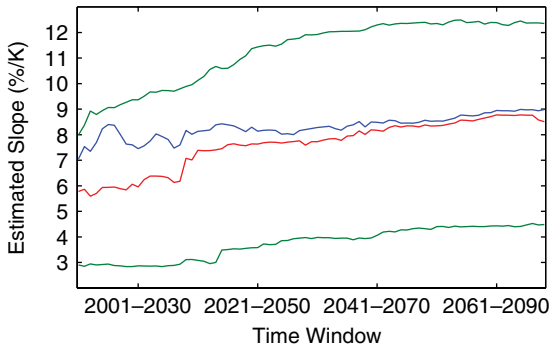


Figure 3. Grid point from the ECHAM5-RCA model with $\beta_1$ slopes (TX90P per Kelvin warming of annual mean daily maximum temperature) estimated with the least-squares (red) and time-slice (blue) approach using the reference period (1961–1990) and a 30-year moving time window. The horizontal axis shows the time period covered by the moving 30-year time window. The green lines represent the range of estimated values for all 12 RCMs.

the internal variability (noise) for the coming decades. However, the signal-to-noise ratio increases with time and, therefore, the estimated slope parameter converges to approximately 9%/K. In addition, the uncertainty in the estimates decreases when the moving window is shifted towards the end of the 21st century, i.e. both curves tend to flatten. For the least-squares approach, for instance, the uncertainty decreases by approximately 75% when the moving time window reaches the end of the 21st century. However, the uncertainty induced by internal variability and the differences in the estimated slope parameters between both pattern scaling approaches are at any time much smaller than the inter-model differences depicted as green lines. Even the largest uncertainty of the estimated $\beta_1$ due to internal variability, found when the time window is close to the reference period, is smaller than 10% of the model range. However, the magnitude of decrease in the uncertainty of the estimated $\beta_1$ exhibits some spatial dependency. The largest decrease in the uncertainty occurs over Scandinavia for three quarters of the GCM–RCM chains in case of the least-squares method and for all RCMs for the time-slice method. For both pattern scaling methods and all GCM–RCM chains, significant decreases in the uncertainty induced by internal variability take place over the Mediterranean.

The aforementioned results are mainly observed for warm extreme indices associated with temperature thresholds in the upper tail of the underlying temperature distribution. For extreme indices with temperature thresholds in the lower tail of the temperature distribution (cold extreme indices) other issues, such as nonlinearities and the convergence of the extreme index values towards zero, have a greater influence on the estimations.

## 5. Discussion

In the last section there is evidence for a violation of the underlying linearity assumption for some of the investigated extreme indices and especially in case of the CWFI. The degree of linearity and the trend magnitude of the extreme index are related to trends in the explanatory variable and the probability of exceeding a particular temperature threshold. Figure 4 illustrates this link in case of the TX10P and TX90P index over the Mediterranean simulated from 1951 to 2099 with the HadCM3Q3-HadRM3Q3. The trend in the yearly mean maximum temperature for the time period from 1951 to 2099, shown in Figure 4(a), can be characterized by a second order polynomial regression model (blue solid line). The fitted curve shows that under the assumption of constant central moments higher than first-order the trend implies a nonuniform shift of the whole temperature distribution towards higher temperatures. The shift itself induces changes in the probabilities of exceeding a particular temperature threshold and therefore in the extreme index. The magnitude and direction of this induced trend depends on the shift direction and the shape of the temperature distribution. Figure 4(c) shows how the probability of exceeding the 90th percentile changes as a function of changing location parameter $\Delta\mu$ in case of a standard normal distribution with three different values for the scale parameter $\sigma$. Figure 4(c) implies that a shift towards higher temperatures induces first a nonlinear increase in the percentage of days per year exceeding the 90th percentile of the reference period but then the trend becomes nearly linear after some point in time. This is exactly what is seen in Figure 4(b) showing the time series of the TX90P and TX10P index. In addition, because the induced nonlinear trend in the time series of the TX90P index resembles the nonlinear trend in the time series of the yearly mean temperature and both time series are in phase, the resulting trend between the extreme index and the explanatory variable is nearly linear. But Figure 4(c) illustrates that the linear trend magnitude can be very different for temperature distributions with the same constant shape but different scale parameter. Further, the figure implies that if changes in the scale parameter occur during the shift of the whole temperature distribution the resulting trend becomes nonlinear.

In contrast to this, changes in the exceedance probabilities at the lower tail of the temperature distribution, illustrated with the TX10P index, exhibit a decrease in the exceedance probability. However, the magnitude of the decreasing trend is much smaller than for the increasing trend in the TX90P index. This explains why the skill values for warm extreme indices (temperature threshold in the upper tail of the temperature distribution) are often much better than for cold extreme indices (thresholds in the lower tail). Furthermore, this explains why the estimated response patterns for the warm indices (i.e. TX90P, TN90P, HWFI and SU) are less affected by the internal variability. As a consequence of this trend, magnitude and linearity dependence on the shift direction of the whole temperature distribution, interpolated estimates for extreme indices (i.e. 2021–2050 based on 1961–1990 and 2070–2099) have higher skill than the extrapolated values. This implies that under the assumption of linearity

 

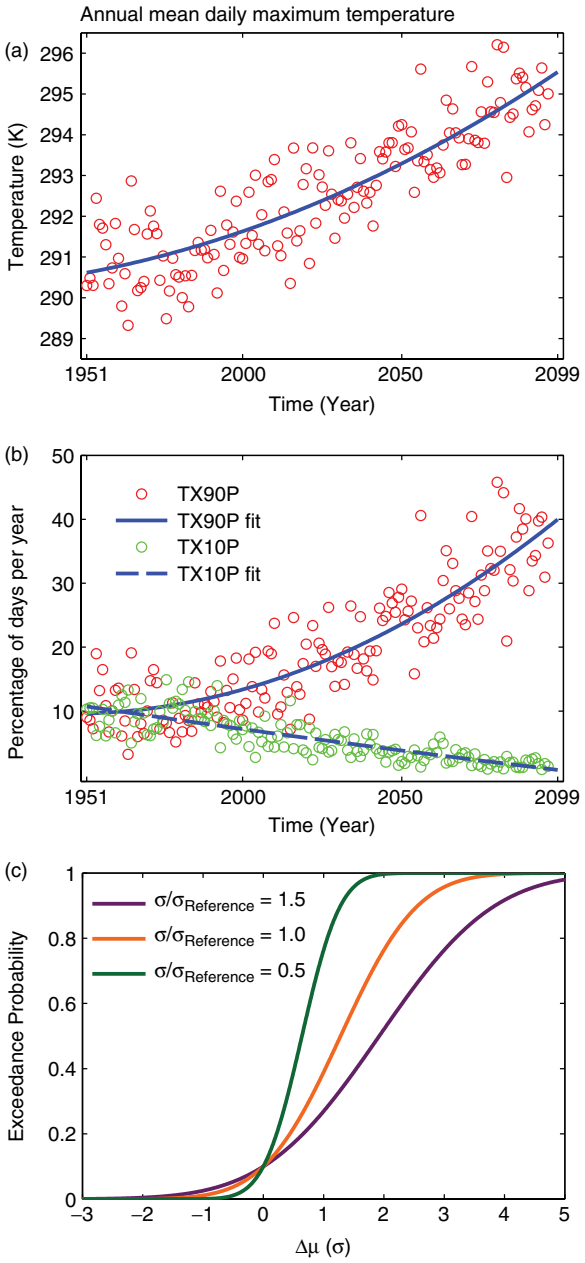


Figure 4. (a) Time series of the annual mean daily maximum temperature (red circles) simulated over the Mediterranean from 1951 to 2099 with the HadCM3Q3-HadRM3Q3 under the SRES A1B emission scenario. The blue solid line represents the second-order polynomial fit. (b) Time series of the TX90P index (red dots) and TX10P (green dots) and the second-order polynomial fits (blue line). (c) Probabilities of exceeding the 90th percentile of a standard normal distribution as a function of changes in the location parameter for three different values of the scale parameter.

there is a dependency of the estimated response pattern on the chosen time periods. The moving time window applied on the TX90P index shows that if the second time slice is placed in the second half of the 21st century the uncertainty of the estimated $\beta_1$ decreases and the estimation converges to a nearly constant value. This effect is especially pronounced over northern Europe, northeast Europe and over the Mediterranean for the TX90P and TN90P index. This corresponds to the regions where the largest warming in the annual mean surface air temperature is expected at the end of the 21st century (Van Der Linden and Mitchell, 2009). For the TN10P index, TX10P index and CWFI, the violation of the linearity assumption is the more dominant issue.

A further issue is that the extreme indices used in this study are so called count data, i.e. they count the number of days or events per time period. Therefore, the range of possible values for the predictands is bounded. In addition, the data is discrete, i.e. for a specific range of yearly mean temperature it is possible to get the same number of days or events per time period. In such cases, simple linear regression models are not appropriate statistical models for dealing with this kind of data. The bounded nature of the data results from the typical property of a cumulative distribution function being bounded between zero and one, as evident in Figure 4(c). The question is whether it is possible to modify the pattern scaling method taking the aforementioned issues into account. There are two potential solutions. First, there is an *ad hoc* approach setting the estimated values to the boundary value when they reach the limit of reasonable values (i.e. no days or all days exceed the temperature threshold). Another possibility is to use a logistic regression model instead of the simple linear regression model (not shown). In general, there is no best alternative regression model or modification of the established simple linear regression model leading to an overall improvement over the whole integration area. The improvement depends on the most dominant issue in each of the simulated trends, i.e. if the nonlinearity or the bounded and/or discrete nature of the data is the main issue. How pronounced these issues are depends on the extreme index and the simulated trends. The magnitude of the trends determines how fast an extreme index reaches a particular boundary value. This is especially pronounced in case of CWFI and HWFI because they have a higher sensitivity on trends in the underlying temperature distribution. As a result, the trend magnitude for both is much higher than for the other extreme indices. All these dependencies explain why the improvements vary between different regions.

For FD and SU, the results suggest that an improvement of the results is possible using seasonal means instead of annual means. This is not surprising because these two indices depend on how pronounced the annual cycle is. The trends in the annual mean temperature over northern Europe are dominated by trends in the DJF mean temperature and over southern Europe by trends in the JJA mean temperature. For this purpose, the largest increase in the skills for FD when using seasonal means are over northern Europe and for SU over the Mediterranean.

## 6. Conclusion

This study demonstrates the potential and limitations of pattern scaling approaches to fill-in gaps in projections of

 

temperature-related extreme indices based on local mean temperature changes. For all extreme indices considered the least-squares approach performs better in interpolating and extrapolating changes in extreme indices from local mean temperature changes than the time-slice approach. The skill for both methods is mainly limited by the degree of linearity of the underlying relationship between mean temperature and index. The linearity of this relationship is determined by the shape of the underlying temperature distribution. In addition, changes in the scale parameter during the shift of the temperature distribution can affect the linearity. However, the shape of the temperature distribution is the main reason why in case of increasing temperatures violations of the linearity assumption occur mainly for cold extremes such as, for instance, the TN10P index or CWFI. As a result, the interpolated and extrapolated estimations have lower skills for cold extreme indices (i.e. TX10P, TN10P, CWFI and FD) than for warm extreme indices (i.e. TX90P, TN90P, HWFI and SU). Given an underlying warming trend, these higher skills are observed for both pattern scaling approaches.

Trends in CWFI are highly sensitive to annual mean temperatures due to the duration minimum of 6 days. Therefore, this extreme index converges faster towards zero than, for instance, the TN10P index. The same problem will occur with ongoing warming for extreme indices characterizing the upper tail of the temperature distribution (e.g. TX90P and HWFI). For the HadCM3Q16-RCA3 model which projects a strong warming, this violation is already evident for the TX90P index over parts of southern Europe at the end of the 21st century.

The results of this study further illustrate that for both pattern scaling methods the skills for the interpolation are better than for the extrapolation. This is to a large extent due to the internal variability that obscures the long-term trend in the near-future period used to derive the extrapolation. Testing the uncertainty induced by internal variability on estimations of $\beta_1$ reveals that estimates of the slope parameter $\beta_1$ are more robust for the second half of the 21st century than for time periods close to the reference period. Setting the second time slice at the end of the 21st century, reductions in the uncertainty of $\beta_1$ up to 75% are observed especially over Scandinavia and the Mediterranean for both pattern scaling approaches. Nevertheless, it is important to note that the model uncertainty exceeds that from internal variability by far.

The results show that simulated trends in the time series of extreme indices can be related to the direction and magnitude of the change in mean temperatures. On the basis of a particular GCM–RCM pair we further demonstrate that the skill of the methods is dependent on regions and seasons considered. Over northern Europe, for instance, the largest projected changes in seasonal mean surface air temperature in case of the SRES A1B scenario occur for the season December–February (Van Der Linden and Mitchell, 2009). On the other hand, for the season June–August the largest projected changes are expected over southern Europe. This implies that annual trends over some regions are dominated by particular seasonal trends. As a consequence of this, over such regions it is possible to improve the skills for the estimated FD and SU values when seasonal means are used instead of annual means. Applying, for instance, the least-squares approach on FD using seasonal DJF mean temperature changes lead to RMSE values up to 50% smaller than for annual values.

Another important limitation of the pattern scaling approaches is that count data is bounded and/or discrete leading to nonphysical or unreasonable estimations such as, for instance, negative number of days per year. All these issues lead to the conclusion that the traditional pattern scaling approaches are only useful when the user knows where over the area of interest the aforementioned issues can occur and how pronounced they are. Alternative pattern scaling techniques such as, for instance, the logistic regression model or *ad hoc* procedures discarding unreasonable values can improve the estimations. But there is no overall solution possible due to the spatial dependency of the discussed issues and the dependency on the GCM–RCM pair.

## References


Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, Klein Tank AMG, Haylock M, Collins D, Trewin B, Rahimzadeh F, Tagipour A, Rupa Kumar K, Revadekar J, Griffiths G, Vincent L, Stephenson DB, Burn J, Aguilar E, Brunet M, Taylor M, New M, Zhai P, Rusticucci M, Vazquez-Aguirre JL. 2006. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research-Atmospheres* **111**: D05109. DOI: 10.1029/2005JD006290.

Huntingford C, Cox PM. 2000. An analogue model to derive additional climate change scenarios from existing GCM simulations. *Climate Dynamics* **16**: 575–586. DOI: 10.1007/s003820000067

Katz RW, Brown BG. 1992. Extreme events in a changing climate: variability is more important than averages. *Climatic Change* **21**: 289–302. DOI: 10.1007/BF00139728

Kendon EJ, Jones RG, Kjellström E, Murphy JM. 2010. Using and designing GCM-RCM ensemble regional climate projections. *Journal of Climate* **23**: 6485–6503. DOI: 10.1175/2010JCLI3502.1

Kennett E, Buonomo E. 2006. Methodologies of pattern scaling across the full range of RT2A GCM ensemble members. Met Office Hadley Centre for Climate Prediction and Research: Exeter, UK.

Mitchell TD. 2003. Pattern scaling - An examination of the accuracy of the technique for describing future climates. *Climatic Change* **60**: 217–242. DOI: 10.1023/A:1026035305597

Mitchell TD, Hulme M. 1999. Predicting regional climate change: living with uncertainty. *Progress in Physical Geography* **23**: 57–78. DOI: 10.1177/030913339902300103

Moberg A, Jones PD, Lister D, Walther A, Brunet M, Jacobeit J, Alexander LV, Della-Marta PM, Luterbacher J, Yiou P, Chen D, Klein Tank AMG, Saladie O, Sigro J, Aguilar E, Alexandersson H, Almarza C, Auer I, Barriendos M, Begert M, Bergstrom H, Bohm R, Butler CJ, Caesar J, Drebs A, Founda D, Gerstengarbe FW, Micela G, Maugeri M, Osterle H, Pandzic K, Petrakis M, Srnec L, Tolasz R, Tuomenvirta H, Werner PC, Linderholm H, Philipp A, Wanner H, Xoplaki E. 2006. Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000. *Journal of Geophysical Research-Atmospheres* **111**: D22106. DOI: 10.1029/2006DJ007103

Plackett RL. 1950. Some theorems in least squares. *Biometrika* **37**: 149–157.

Ruosteenoja K, Tuomenvirta H, Jylha K. 2007. 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. DOI: 10.1007/s10584-006-9222-3

Simolo C, Brunetti M, Maugeri M, Nanni T, Speranza A. 2010. Understanding climate change in- duced variations in daily temperature distributions over Italy. *Geophysical Research Letters* **115**: D22110. DOI: 0.1029/2010JD014088

Toreti A, Desiato F. 2008. Changes in temperature extremes over Italy in the last 44 years. *International Journal of Climatology* **28**: 733–745. DOI: 10.1002/joc.1576

Van der Linden P, Mitchell JFB. 2009. ENSEMBLES: climate change and its impacts: summary of research and results from the ENSEMBLES project. Met Office Hadley Centre: Exeter, UK.

 