Vol. 48: 163–176, 2011
doi: 10.3354/cr00906

CLIMATE RESEARCH
Clim Res

Published August 30

*Contribution to CR Special 27 'Climate change in the NW Iberian Peninsula'*


# Statistical downscaling of daily temperatures in the NW Iberian Peninsula from global climate models: validation and future scenarios

**S. Brands[1,2,*], J. J. Taboada[2], A. S. Cofiño[3], T. Sauter[4], C. Schneider[4]**

[1]Instituto de Física de Cantabria (CSIC-UC), 39005 Santander, Spain
[2]MeteoGalicia, Xunta de Galicia, 15707 Santiago de Compostela, Spain
[3]Department of Applied Mathematics and Computer Science, Universidad de Cantabria, 39005 Santander, Spain
[4]Department of Geography, RWTH Aachen University, 52056 Aachen, Germany

ABSTRACT: We used the analogue method to generate ensemble projections of local daily mean, maximum and minimum air temperatures in the NW Iberian Peninsula until the middle of this century. A 3-step method was followed. (1) The error of the analogue method under optimal conditions was estimated, using air temperatures at 850 hPa and mean sea level pressure from reanalysis data as predictor variables. (2) The method's error under suboptimal conditions was assessed by taking these predictors from control runs of a multi-model, multi-initial-conditions ensemble of global climate models. Neither the predictor data nor the downscaled series were corrected. Under these suboptimal conditions, none of the individual downscaled series could robustly reproduce the cumulative distribution function (CDF) of the observations in any season of the year. However, when the single downscaled series were combined into a multi-model series, CDFs were reliably reconstructed for summer and autumn. (3) Temperature series were downscaled from the ensemble's scenario runs and compared to observations in the reference period to detect local climate change. In addition to the mean relative warming, it can be shown that the less frequent the event in the reference period, the higher its frequency increase and the broader its uncertainty interval in the scenario period. This tendency is more pronounced for daytime than for night-time heat/warm events, leading to a tripling to quadrupling of the former in summer. The local projections' uncertainty intervals are dominated by model errors rather than by forcing or initial-conditions uncertainties.

KEY WORDS: Statistical downscaling · Global climate models · GCM · Multi-model · Climate projections · Uncertainty · Air temperature · Extreme events · Iberian Peninsula



## 1. INTRODUCTION

> 'It may well happen that different GCMs produce different results' (Von Storch et al. 1993, p. 1170).

Since von Storch et al. (1993) wrote that line, the statistical downscaling (SD) community has undertaken much effort to generate regional to local climate projections from (coupled) global climate models (GCMs). However, most studies are based on a single GCM (Hewitson & Crane 1996, Wilby et al. 1999, Huth 2004, Frías et al. 2006, Schmidli et al. 2007) and therefore do not take into account the significant 'model uncertainty' (Stainforth et al. 2007) of multi-model ensembles (Christensen et al. 2007). Ignoring model uncertainty means ignoring an important contributor to the total uncertainty of climate projections (Stainforth et al. 2007, Knutti et al. 2010) and might lead policy-makers, adaption strategists and the general public to the erroneous belief that deterministic climate projections have a prognostic skill (Oreskes et al. 1994).

Regional climate change is usually quantified with the 'delta change method' (Räisänen 2007): future projections are relative to the simulated climate in the control period (Christensen et al. 2007). In this paradigm, one completely remains in the model world and,

*Email: brandssf@unican.es

strictly speaking, it is not necessary to validate the downscaled control climate against observations.

Alternatively, projections are corrected by the error of the downscaled control climate and then presented relative to observed values in the reference period (Imbert & Benestad 2005). However, if a control run cannot reproduce key elements of the observed regional climate, artificial feedback processes in the corresponding scenario run amplify the error, leading to unrealistic projections (Räisänen 2007). No kind of *a posteriori* correction (Karl et al. 1990, Von Storch 1999, Imbert & Benestad 2005) can account for this additional error.

The number of GCM ensemble members used in downscaling studies is mainly determined by the availability, inter-model comparability and quality of the GCM (predictor) data.

On a monthly time scale, the Intergovernmental Panel on Climate Change (IPCC) 4th Assessment Report (AR4) GCM data (Christensen et al. 2007, Randall et al. 2007) is well organised and readily available (PCMDI 2009a), thereby permitting SD studies based on a large multi-model ensemble (Benestad 2005). However, it is the frequency increase of extreme events on a daily timescale which is most important for impact studies and the corresponding adaption strategies (Easterling et al. 2000).

Projections on a daily time scale are therefore needed (Wilby et al. 2004). However, data harmony between the different GCMs of the main climate-research institutes cannot be assumed, although the Coupled Model Intercomparison Project 5 (CMIP5) (PCMDI 2009b) may offer improvement in this area. Harmonising daily predictor data from different GCMs for use in a multi-model ensemble is a time-consuming process. This might explain why there are so few SD studies based on such ensembles; to our knowledge, only Timbal et al. (2003), Timbal & Jones (2008) and Teutschbein et al. (2011) have used daily predictor fields from a small multi-model ensemble.

On the other hand, the additional uncertainty introduced by the various SD techniques seems to have been sufficiently investigated (see Fowler et al. 2007 and references therein) as it was the main subject of the Statistical and Regional Dynamical Downscaling of Extremes for European Regions (STARDEX) project (Goodess 2005).

In the present study, air temperatures at the 850 hPa pressure level and mean sea level pressure (T850MSL) were taken as the optimal predictor combination to downscale local daily mean (*T*mean), maximum (*T*max) and minimum (*T*min) air temperatures at 2 m height. This choice is based on a detailed predictor screening using re-analysis data in the validation period (optimal conditions). The corresponding results were consistent with those of Timbal et al. (2003) and therefore are not shown in the present article. For downscaling of *T*min, accuracy is enhanced if humidity predictors are added to temperature and circulation predictors (Timbal et al. 2003). However, we assume that humidity variables are not reliably reproduced by the current GCM generation (Timbal et al. 2003, Maraun et al. 2010), and therefore we did not use them in the present study.

As for the SD technique, we applied the analogue method (AN) (Lorenz 1969, Zorita et al. 1995, Zorita & Von Storch 1999). Inspired by Wilby et al. (2004), our study is outlined as follows.

In Section 2 the predictor and predictand data are presented. Section 3 describes the downscaling method and the validation measures. Section 4 presents the results of a 3-step downscaling procedure (see Table 1). In Section 4.1 (Step 1), we estimated the SD technique's error by taking the predictors in the validation period from re-analysis data (optimal conditions). In Section 4.2 (Step 2), predictors in the validation period were taken from control runs of an ensemble of AR4 GCMs (Solomon et al. 2007); the additional model uncertainty was taken into account (Chen et al. 2006) and thus downscaling conditions were suboptimal. In Section 4.3 (Step 3), predictands were downscaled from the ensembles' scenario runs. In contrast to Wilby et al. (2004), the resulting local climate projections are presented relative to observations in the reference period. No form of *a posteriori* corrections (Karl et al. 1990, Von Storch 1999, Imbert & Benestad 2005), either of the predictors or of the downscaled series, was conducted. Table 1 gives detailed information on the test and validation periods used in the 3 downscaling steps. Section 5 discusses the main results and Section 6 provides some concluding remarks.

The present study focused on the summer (JJA) and autumn (SON) seasons in the NW Iberian Peninsula

Table 1. The 3-step downscaling procedure followed in the present study. 'Section' refers to section of text in the present paper. GCM: global climate model

| Downscaling step | Section | Training period | Validation and projection period |
|---|---|---|---|
| 1: Optimal conditions | 4.1 | 1 Sep 1972 to 31 Aug 1991 Re-analysis data | 1 Sep 1992 to 31 Aug 2002 Re-analysis data |
| 2: Suboptimal conditions | 4.2 | 1 Sep 1972 to 31 Aug 2002 Re-analysis data | 1 Sep 1972 to 31 Aug 1999 GCM control-run data |
| 3: Projections | 4.3 | 1 Sep 1972 to 31 Aug 2002 Re-analysis data | 1 Sep 2002 to 31 Aug 2050 GCM scenario-run data |

(Galicia) (Lorenzo et al. 2008). The presented 3-step method can be applied in any region where quality-controlled long-term observations are available.

## 2. DATA

Two classes of daily predictor data were used:

(1) The European Centre for Medium-Range Weather Forecasts (ECMWF) 40-yr re-analysis data (ERA-40) (Uppala et al. 2005), downloaded from the public server of the ECMWF (2009a); and

(2) GCM data from 5 control runs and 14 scenario runs of 3 different GCMs. The models were taken from the ENSEMBLES-stream 1 project (Niehörster et al. 2008) and are members of the AR4 multi-model ensemble (Randall et al. 2007). Control-run data were used to validate the AN under suboptimal conditions (Step 2), while scenario-run data were used to generate climate projections (Step 3). The GCMs were obtained from the CERA database (http://cera-www.dkrz.de/CERA/) and were post-processed on the supercomputer of the Deutsche Klimarechenzentrum (http://cera-www.dkrz.de/CERA).

The Gaussian grids of the GCM data were regridded on the regular ERA-40 2.5° grid using the bicubic interpolation technique (Jones 1999). To circumvent the problem of different time aggregations in GCM and reanalysis data, daily means were calculated for all predictor data. Furthermore, different variable names and units were harmonised to correspond to the ECMWF standard table 128 (ECMWF 2009b). The chosen domain (Fig. 1a) was kept constant throughout the work and was assumed to capture the 'skilful scale' (Benestad et al. 2008) of the 3 GCMs.

In the first step (Section 4.1), T850MSL predictor data were taken from ERA-40 for both the validation (1 Sep 1972 to 31 Aug 1991) and training (1 Sep 1992 to 31 Aug 2002) periods (Table 1); the AN was validated under optimal conditions.

In the second step (Section 4.2), T850MSL predictor data were taken from ERA-40 and from 5 control runs of the GCM experiments listed under Step 2 in Table 2. The ERA-40 data served as predictors in the training period (1 Sep 1972 to 31 Aug 2002) while the GCM data were used in the validation period (1 Sep 1972–31 Aug 1999, control-run days). Thus, the conditions for validating the AN were suboptimal in Step 2. As both the model uncertainties (3 different GCMs) and initial-conditions uncertainties (3 runs of the MPI-ECHAM5 GCM) were described by the 5 control runs, we hereinafter use the term 'multi-model multi-initial conditions ensemble' (MMMICE) (Smith 2004) for this.

In the third step (Section 4.3), predictor data in the projection period (1 Sep 2002 to 31 Aug 2050) came from 14 scenario runs of the MMMICE, spanning the Special Report on Emissions Scenarios (SRES) emission scenarios A1B, A2 and B1 (Nakicenovic & Swart 2000) (Step 3 in Table 2). Thus, in addition to model and initial-conditions uncertainties, forcing uncertainties were taken into account. The predictors used in the training period are identical to those of Step 2.

As predictand data, daily *T*mean, *T*max and *T*min at 9 synoptic stations of the Spanish and Galician meteorological services (Agencia Estatal de Meteorología and MeteoGalicia) were used, covering both the coastal and more continental regions of the area under study (Fig. 1b). These time series were quality-controlled and homogenised as described in Cruz et al. (2009).

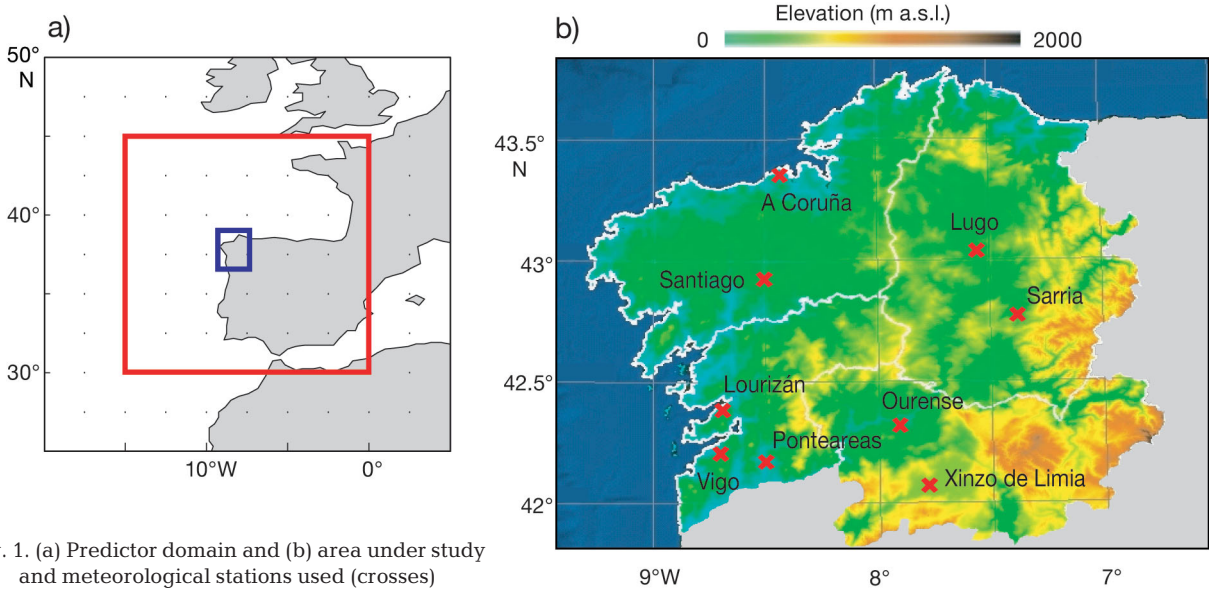


Fig. 1. (a) Predictor domain and (b) area under study and meteorological stations used (crosses)

Table 2. General circulation model (GCM) data used in the validation and projection period of Steps 2 and 3

| Step | GCM | Scenario (run) | Acronym | Source |
|---|---|---|---|---|
| 2 | MPI-ECHAM5 | Control (1, 2, 3) | M1, M2, M3 | Roeckner (2007a) |
| | CNRM-CM3 | Control (1) | C1 | Royer (2006a) |
| | NERSC, bccr_bcm2 | Control (1) | B1 | Drange (2006a) |
| 3 | MPI-ECHAM5 | A1B (1, 2, 3) | M1A1B, M2A1B, M3A1B | Roeckner (2007b) |
| | | A2 (1, 2, 3) | M1A2, M2A2, M3A2 | Roeckner (2007c) |
| | | B1 (1, 2, 3) | M1B1, M2B1, M3B1 | Roeckner (2007d) |
| | CNCM3 | A1B (1) | C1A1B | Royer (2006b) |
| | | A2 (1) | C1A2 | Royer (2006c) |
| | | B1 (1) | C1B1 | Royer (2006d) |
| | NERSC, bccr_bcm2 | A1B (1) | B1A1B | Drange (2006b) |
| | | A2 (1) | B1A2 | Drange (2006c) |

## 3. METHODS

### 3.1. The analogue method

The analogue method, introduced to the atmospheric sciences by Lorenz (1963, 1969), is a widely used SD technique. As the efficiency of the AN decreases with increasing dimensionality of the predictor fields (Van den Dool 1994), Barnett & Preisendorfer (1978) applied a principal components analysis (PCA; see Preisendorfer & Mobley 1988 for the application of PCA in meteorology) to reduce the dimensions of the data. They then defined the axes of the state space that the analogues are searched in using $k$ principal component vectors (PCs) instead of $v$ raw atmospheric variables. As $k < v$, the dimensionality of the predictor data is drastically reduced if the variable is spatially autocorrelated (redundant). This procedure was followed by Zorita et al. (1995), who generated precipitation scenarios for the coastal USA, and by Cubasch et al. (1996), who projected future temperature and precipitation scenarios in southern Europe. Working in optimal conditions, Zorita & Von Storch (1999) compared it with other linear and nonlinear SD techniques, and concluded:

> 'It is found in these applications that the analog method performs in general as well as the more complicated methods, and can be applied to both normally and nonnormally distributed local variables' (Zorita & Von Storch 1999, p. 2474).

In this classical or in slightly modified forms, the AN has recently been used in weather (Gutiérrez et al. 2004) and seasonal forecasting (Sordo et al. 2008). For regional to local climate change studies, it is quite appropriate (Timbal & McAvaney 2001, Timbal et al. 2003, Imbert & Benestad 2005, Frías et al. 2006, Timbal & Jones 2008) as it reproduces the predictand's marginal probability density function (PDF) fairly well, at least in optimal conditions (Zorita & Von Storch 1999).

However, the classical version of the AN (Zorita & Von Storch 1999) is not able to simulate record events in a warming climate because the analogue search is restricted to the target day's season of the year (restricted AN). To overcome this, Imbert & Benestad (2005) suggested expanding the analogue search to all seasons (unrestricted AN). Yet, this only partly solves the problem, as no record heat events can be projected in summer (Imbert & Benestad 2005).

Finally, if the autocorrelation and/or extreme value distribution of the local predictand are downscaled, the nearest-neighbour resampling method (Lall & Sharma 1996) yields better results than the AN (Brandsma & Buishand 1998).

In the present study, the following variation of the classical AN (Zorita & Von Storch 1999) was applied in each of the 3 working steps.

First, an S-mode PCA (Huth 1996) was conducted to reduce the dimensions of the predictor data. Let $X_{tv}$ be a matrix with $t$ time steps and $v$ predictor variables ($v$ = number of atmospheric variables × number of horizontal grid points × number of pressure or surface levels). After standardising the predictor variables column by column, they were geographically weighted by multiplying them with the absolute value of the cosine of their corresponding latitude (Benestad et al. 2008). On the basis of the resulting matrix $X'_{tv}$, the covariance matrix $K_{VV}$ was calculated:

$$K_{vv} = \frac{1}{v-1} X'^{T}_{tv} \; X'_{tv} \quad (1)$$

where $X'^{\mathrm{T}}_{tv}$ is the transpose of $X'_{tv}$, and $v - 1$ is the spatial degrees of freedom. This then serves to extract the eigenvectors and eigenvalues by applying the singular value decomposition algorithm (Wilks 2006, Press et al. 2007):

$$K_{vv} = L_{vv}\, \Omega\, R_{vv}{}^{T} \quad (2)$$

where $L_{vv}$ is a square matrix with $v \times v$ dimensions and in which columns = eigenvectors = empirical orthogo-

nal functions (EOFs); Ω is a diagonal matrix with $w = v$ eigenvalues in descending order; and $R_{vv}^{T}$ is a transposed square matrix with $v \times v$ dimensions. The projection of the EOFs in $L_{vv}$ on the standardised and geographically weighted matrix $X'_{tv}$ was achieved by the following matrix multiplication:

$$C_{tk} = X'_{tv} L_{vv} \quad (3)$$

where $C_{tk}$ is a matrix with $t$ time step, and $k$ PCs. In the resulting matrix $C_{tk}$, the PCs are ordered from left to right in descending order of their explained variance, which is calculated by dividing the corresponding eigenvalues in (Eq. 2) by the number of input atmospheric variables $v$. The $k$ PCs yielding an explained variance >99% were chosen to describe the axes of the state space. They were weighted by their eigenvalue as proposed by Imbert & Benestad (2005).

In Fig. S1 in the Supplement (available at www.int-res.com/articles/suppl/c048p163_supp.pdf), the results of the PCA for individual ERA-40 predictors (period: 1 Sep 1972 to 31 Aug 2002) are shown. The acronyms follow Table 128 of ECMWF (2009b). Fig. S2 in the supplement shows EOF 1 to 9 for the mean sea level pressure (MSL).

To reproduce the dispersion of the local variable to predict, only one predictand value per target day, corresponding to the nearest neighbour, was assumed. The nearest neighbour is defined by minimising the Euclidean distance, which is commonly used in this context (Matulla et al. 2008). It is important to mention that least-squares validation measures like the correlation coefficient can be improved by calculating the (distance-weighted) mean of various neighbour values (Timbal & McAvaney 2001), with the disadvantage of underestimating the predictand's dispersion. As suggested by Imbert & Benestad (2005) the unrestricted AN was applied.

In the first step (optimal conditions), $X_{tv}$ consisted entirely of ERA-40 data and was split into a training period of 19 yr (1 Sep 1972 to 31 Aug 1991; Table 1) and a validation period of 10 yr (1 Sep 1992 to 31 Aug 2002). As 2 predictor variables (T850MSL) entered the PCA, the procedure can be called combined PCA (Bretherton et al. 1992).

In the second step, a common PCA (Benestad 2002) was applied: ERA-40 data constitutes the upper part of $X_{tv}$ (training period: 1 Sept 1972 to 31 Aug 2002) and GCM control-run data its lower part (validation period: 1 Sep 1972 to 31 Aug 1999, control-run days; Table 2). The control-run data were assumed to be independent, as they do not covary with their re-analysis counterparts. Standardisation, weighting and PCA were conducted on this 'common' matrix.

In the third step, the procedure of Step 2 was repeated, with the difference that scenario-run data was used instead of control-run data (Table 3). The projection period extends from 1 Sep 2002 to 31 Aug 2050. Finally, a scenario period (SCE) was defined for JJA 2020–2050 and SON 2020–2049, and the corresponding downscaled temperature scenarios were compared to observations in the reference period (REF: JJA 1973–2002 and SON 1972–2001, respectively).

### 3.2. Validation measures

To compare the cumulative distribution functions (CDFs) of the observed and downscaled time series, the reliability score described in Perkins et al. (2007) (which we will call PRS) was calculated. For a perfect fit of both CDFs, PRS equals 1, but is 0 in the case of no overlap.

In addition, we concentrated on certain 'regions' of the CDF using the following validation measures recommended by the STARDEX project (Goodess 2005): the bias (BIAS) and the difference between the forecasted and observed 10th and 90th percentiles (DPCT10 and DPCT90, respectively):

$$\mathrm{BIAS} = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i) \quad (4)$$

where $n$ is the number of time steps (days); $o_i$ is the $i$th value of the observations; and $f_i$ is the $i$th value of the forecast.

$$\mathrm{DPCT10} = \mathrm{PCT10}_f - \mathrm{PCT10}_o \quad (5)$$

$$\mathrm{DPCT90} = \mathrm{PCT90}_f - \mathrm{PCT90}_o \quad (6)$$

where $\mathrm{PCT10}_f$ and $\mathrm{PCT90}_f$ are the 10th and 90th percentiles of the forecast, respectively; and $\mathrm{PCT10}_o$ and $\mathrm{PCT90}_o$ are the 10th and 90th percentiles of the observations, respectively. These validation measures are usually given in °C, not taking into account that an

Table 3. Definition of heat and warm events. The units for all events are % of reference period (REF), with 0% meaning no increase. *T*max and *T*min: daily maximum and minimum air temperatures, respectively; PCT90, PCT95 and PCT97.5: 90th, 95th and 97.5th percentiles, respectively

| Type<br>Name of event | Definition |
|---|---|
| **Daytime heat events** | |
| Warm days | No. of days with *T*max > PCT90 |
| Heat days | No. of days with *T*max > PCT95 |
| Extreme heat days | No. of days with *T*max > PCT97.5 |
| **Night-time warm events** | |
| Mild nights | No. of days with *T*min > PCT90 |
| Warm nights | No. of days with *T*min > PCT95 |
| Extreme warm nights | No. of days with *T*min > PCT97.5 |

error of 1°C is more important for a less dispersed time series (e.g. summer *T*min at A Coruña) than for a more dispersed one (e.g. autumn *T*max at Ourense). Here, both the observed and downscaled time series were divided by an outlier-resistant dispersion measure of the observations before entering the validation: the mean absolute deviation (MAD) from the median (Press et al. 2007):

$$\mathrm{MAD} = \frac{1}{n}\sum_{i=1}^{n}(o_i - \mathrm{MED}_o) \quad (7)$$

where $n$ is the number of time steps (days); $o_i$ is the $i$th value of the observations; and $\mathrm{MED}_o$ is the median of the observations.This procedure permits a comparison of the validation results for various sites, seasons and predictands.

Further, the 2-sided *t*-test for significant bias, modified for serially correlated data (Wilks 2006), is shown in binary format (see SIGTEST in Figs. 2 to 5 and Fig. S3 in the supplement): it equals 1 if the bias is significant at the 0.1% level and 0 if the $H_0$ hypothesis of equal means cannot be rejected. The test statistic ($z$) of this parametric hypothesis test is given by:

$$z = \frac{\mathrm{BIAS} - \mu}{\frac{s_d^2}{n}} \quad (8)$$

where μ is the mean of the theoretical Gaussian distribution, and $s_d^2/n$ is the variance of the differences of $n$ paired values (= observation and forecast). For $H_0$ (= no significant bias), μ equals 0 in Eq. (8). The serial correlation of daily temperature data artificially increases the $z$-value, which leads to an erroneous acceptance of the alternative hypothesis. Therefore the effective sample size ($n^*$) replaces $n$ in Eq. (7):

$$n^* = n\frac{1 - \mathrm{R\ LAG1}}{1 + \mathrm{R\ LAG1}} \quad (9)$$

where RLAG1 is the lag-1 (Wilks 2006) autocorrelation coefficient. In Steps 1 and 2, error bars were calculated by the nonparametric bootstrap percentile method (Efron & Tibshirani 1993). Based on 1000 synthetic pairs of observed and downscaled time series, the 95% confidence interval of the validation measure's $H_0$ was estimated (Wilks 2006).

In Step 3, the error bars represent the standard error of $n$ values of the respective statistical parameter. These $n$ values were calculated on $n$ downscaled series per station, predictand and season of the year. Here, $n$ is defined by the number of scenario runs, and is equal to 14. As the definition of uncertainty intervals for climate change projections is subject to active scientific debate (Stainforth et al. 2007), we do not want to introduce an additional uncertainty source for defining these uncertainties (Katz 2002). As a consequence, the bootstrap percentile method was not applied in this step.

CDFs were calculated by nonparametric kernel density smoothing (Wilks 2006). Gaussian kernels were assumed and the (modelled) cumulative probability was calculated for a class width of 0.1°C, ranging from the minimum to maximum value of the observed or downscaled time series respectively.

## 4. RESULTS

### 4.1. Step 1: validation under optimal conditions

Fig. 2 shows PRS, BIAS, DPCT10, DPCT90 and SIGTEST for the summer values of *T*mean, *T*max and *T*min. On the *x*-axis, the equivalent of 1 MAD in °C is given for each station.

As shown by the PRS scores being close to 1, the predictands' CDFs are reproduced well by the AN, using T850 and MSL as predictors (AN+T850MSL). Admittedly, the downscaled series tend to significantly underestimate the local mean temperatures. However, this cold bias was only slight at most of the stations, while the upper and lower percentiles were not reproduced significantly worse than the mean (compare PCT10 and PCT90 with BIAS in Fig. 2). This confirms that the probability of relatively rare heat/cold events is reproduced well by AN+T850MSL when conditions are optimal. In addition, the accuracy (not shown) was comparable to the restricted AN (Timbal & McAvaney 2001, Timbal et al. 2003).

In autumn, the reliability of the downscaled series was slightly higher than in summer (Fig. S3 in the supplement), while in winter, comparable results were yielded (not shown). Comparing spring to summer results, *T*mean and *T*max were worse reproduced while *T*min was more reliably downscaled (not shown).

### 4.2. Step 2: validation under suboptimal conditions

In Step 2, we validated whether AN+T850MSL was able to reproduce the CDFs of local daily observations when predictor data in the validation period were taken from GCM control runs (Step 2 in Table 2). Figs. 3 to 5 and Figs. S4 & S5 in the supplement show the combined error of the AN and the control run the predictor data are taken from.

As the predictors were extracted from 5 control runs (Table 2), 5 downscaled series were generated for each station, season of the year and predictand (B1 = NERSC, BCCR_BCM2 run 1, C1 = CNRM-CM3 run 1, M1 to M3 = MPI-ECHAM5 run 1 to 3; Table 2). By simply joining these, a single multi-model sample (MM) was created for each site, season and predictand.

First, it has to be pointed out that none of the individual downscaled series can robustly reproduce the local

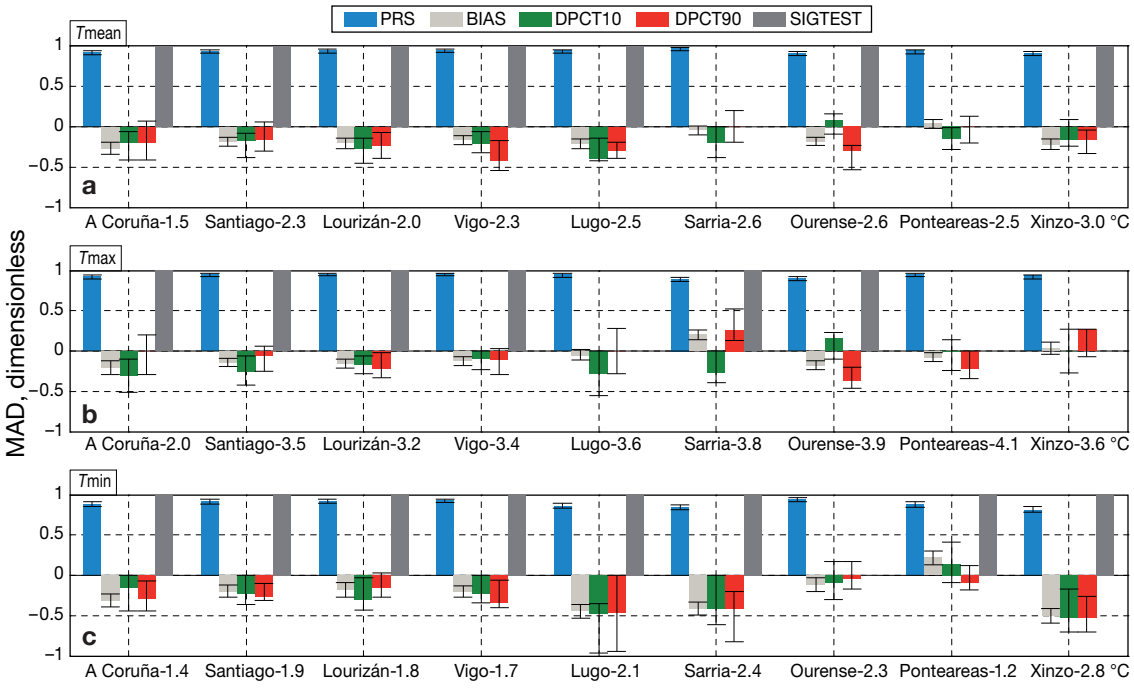


Fig. 2. Validation under optimal conditions in summer (JJA) for daily (a) mean (*T*mean), (b) maximum (*T*max) and (c) minimum (*T*min) air temperatures. MAD: mean absolute deviation from the median; PRS: reliability score described in Perkins et al. (2007); BIAS: the bias described in Eq. (4); DPCT10 and DPCT90: difference between forecasted and observed 10th and 90th percentiles, respectively; SIGTEST: 2-sided *t*-test for significant bias; error bars: 95% confidence intervals, obtained using the bootstrap percentile method. On the *x*-axis, the equivalent of 1 MAD in °C is given for each station

observations for all predictands and stations in any season of the year. With up to –1 MAD (–4°C), the bias was most evident for *T*max in spring downscaled from the NERSC, BCCR_BCM2 control run (not shown).

In contrast, calculating the CDF from MM yielded a robust good fit for all predictands and stations in summer and autumn.

This most important result is shown for the summer season in A Coruña (maritime climate) and Ourense (more continental climate) in Figs. 3 to 5.

The left panels of Figs. 3 to 5 show the observed (OBS) and downscaled (B1, C1, M1, M2, M3, MM) CDFs. The corresponding value of 1 MAD in °C is given in the upper left of the subplots.

In the right-hand panels of Figs. 3 to 5, the validation measures (*y*-axis) presented in Section 3.2 are shown for the 5 single-model hindcasts (B1, C1, M1, M2, M3) and the MM hindcast (*x*-axis). The corresponding results for *T*max at all stations are shown in Fig. S4 in the Supplement. It is equally valid for *T*mean and *T*min (not shown). Autumn results were comparable and are shown for *T*max in Fig. S5 in the supplement.

It has to be underlined that neither the predictor data nor the downscaled series were corrected by bias adjusting, variance inflation (Karl et al. 1990), randomisation (Von Storch 1999) or any other kind of correction. The downscaled series therefore are consistent with their corresponding GCM predictor data.

### 4.3. Step 3: projections

In the last step of the present study, ensemble projections downscaled from the scenario runs of the MMMICE (Table 2) were generated for summer and autumn. CDFs of the scenario period (SCE: JJA 2021–2050, SON 2020–2049) were compared to CDFs of the observations in the reference period (REF: JJA 1973–2002, SON 1972–2001). Again, neither the predictors nor the downscaled projections were corrected (Trigo & Palutikof 1999).

Like in Step 2, MM samples were generated for each site, season and predictand by simply joining the 14 corresponding downscaled time series. Due to their good fit in the control period (Section 4.2), climate-change indicators calculated from MM were assumed to provide the most credible projections.

Figs. 6 & 7 show the CDFs of the single downscaled series and the MM samples for the summer *T*max at a maritime site (A Coruña) and a continental site (Ourense), respectively. The GCM acronyms used in Section 4.2 are accompanied by their corresponding SRES scenario (Nakicenovic & Swart 2000). To give an example: 'M3A1B' refers to the third run of the MPI-ECHAM5 driven by the A1B emission scenario.

The ensemble spread of the local projections was clearly dominated by model uncertainties, rather than by forcing or initial-conditions uncertainties (repre-

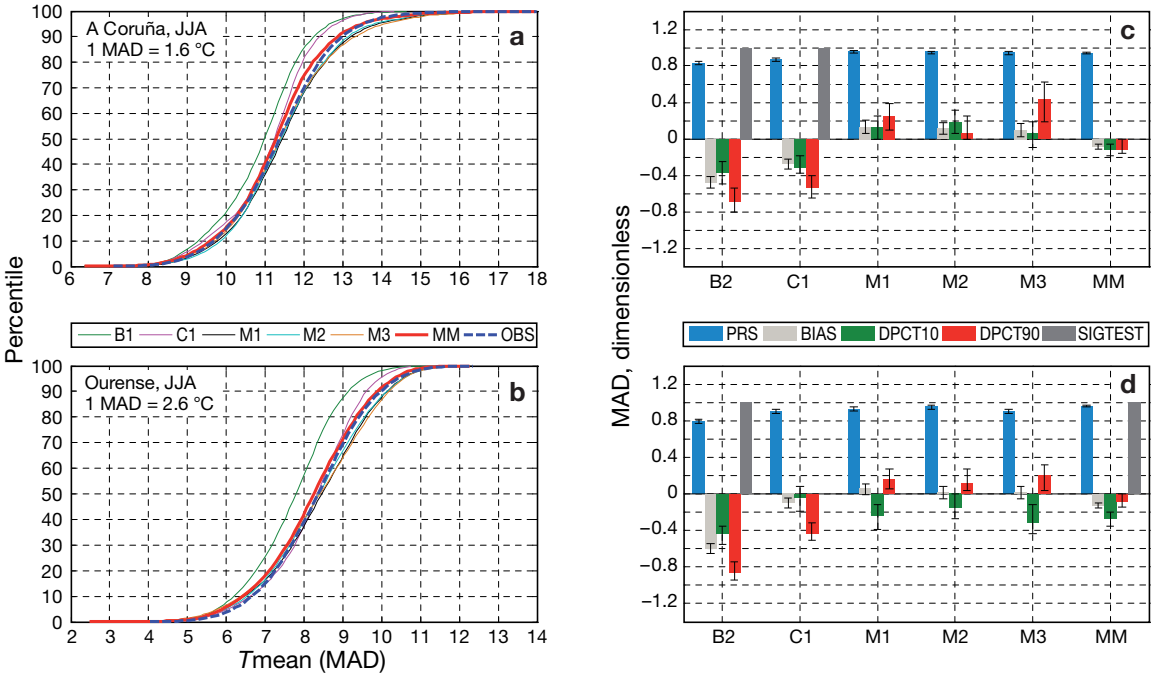


Fig. 3. Validation under suboptimal conditions for daily mean air temperatures (*T*mean) in summer (JJA). Observed (OBS) and downscaled (B1, C1, M1, M2 and M3: see Table 2; MM = multi-model) marginal cumulative distribution functions (CDFs) for (a) A Coruña and (b) Ourense. Validation measures for the 5 single-model hindcasts (B1, C1, M1, M2, M3) and the MM hindcast for (c) A Coruña and (d) Ourense. MAD: mean absolute deviation from the median; PRS: reliability score described in Perkins et al. (2007); BIAS: the bias described in Eq. (4); DPCT10 and DPCT90: difference between forecasted and observed 10th and 90th percentiles, respectively; SIGTEST: 2-sided *t*-test for significant bias; error bars: 95% confidence intervals, obtained using the bootstrap percentile method

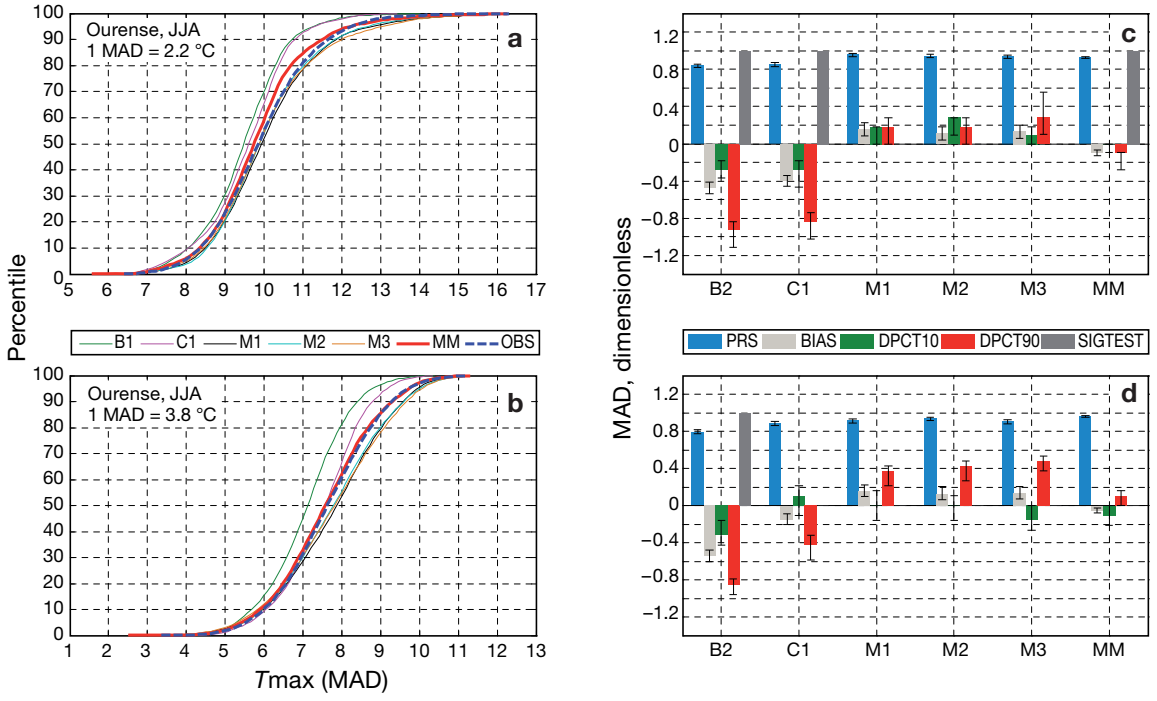


Fig. 4. Validation under suboptimal conditions for daily maximum air temperatures (*T*max) in summer (JJA). See caption to Fig. 3 for panel descriptions and abbreviations

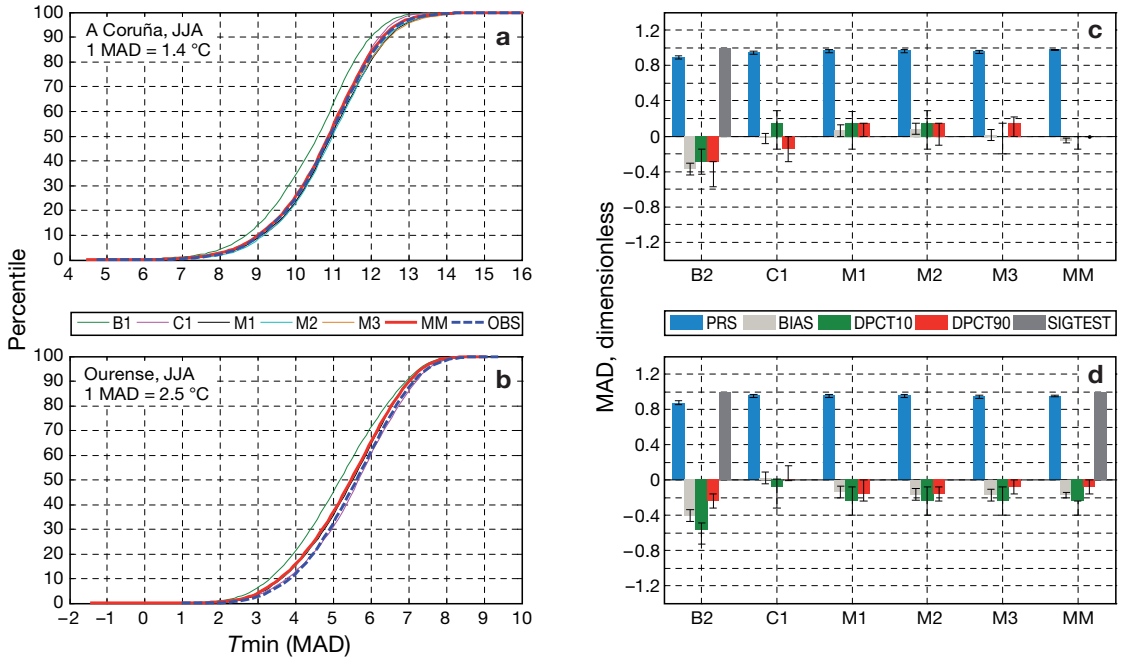


Fig. 5. Validation under suboptimal conditions for daily minimum air temperatures (*T*min) in summer (JJA). See caption to Fig. 3 for panel descriptions and abbreviations

sented by the 3 MPI-ECHAM5 runs), the latter 2 having a similar magnitude (Figs. 6 & 7).

For *T*max (Figs. 6 & 7) and *T*mean (not shown), relative warming was especially pronounced for rare events (as defined by percentile thresholds), a result that was less obvious for *T*min (not shown). Furthermore, for the summer *T*max at A Coruña, the relative warming of all percentiles was less pronounced than in Ourense (compare Figs. 6 & 7). This means that the buffering effect of the sea surface's heat capacity on the relative warming at coastal sites, especially pronounced for *T*max, is reproduced well.

One may ask why the CDFs of the downscaled series based on the NERSC, BCCR_BCM2-scenario runs (B1A1B and B1A2 in Figs. 6 & 7) were colder than the observations' CDFs. This is because the large cold bias in the control period (see green lines for B1 in Figs. 3 to 5 and Figs. S4 & S5 in the supplement) leads to projections in the SCE that remain colder than the observations in the REF.

More specific results for summer and autumn are given in Fig. 8. Fig. 8a,b shows the mean relative warming (MRW) for *T*mean, *T*max and *T*min at each of the 9 stations for both summer (Fig. 8a) and autumn (Fig. 8b). Bars refer to the MRW calculated from MM, and error bars to the standard error of 14 MRW values, calculated from the 14 single downscaled series.

MRW was more pronounced in summer than in autumn, highest for *T*max and lowest for *T*min. The greater the warming, the larger was its corresponding uncertainty interval.

The greatest MRWs and widest uncertainty intervals were simulated for the summer *T*max. They ranged from 0.7 ± 0.2°C to 1.8 ± 0.3°C, followed by *T*mean from 0.5 ± 0.1°C to 1.2° ± 0.2°C, and *T*min from 0.3 ± 0.1°C to 0.8 ± 0.2°C.

Autumn values ranged from 0.0 ± 0.1°C to 1.3 ± 0.2°C for *T*max, from –0.1 ± 0.1°C to 0.8 ± 0.2°C for *T*mean and from –0.1 ± 0.1°C to 0.4 ± 0.1°C for *T*min. For both

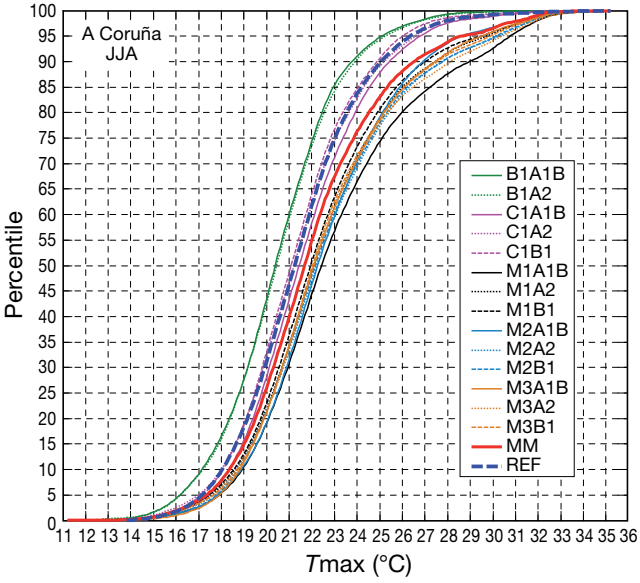


Fig. 6. Projections for daily maximum air temperatures (*T*max) in summer (JJA) at A Coruña (reference period [REF]: JJA 1973–2002; scenario period [SCE]: JJA 2021–2050). MM: multi-model. See Table 2 for other abbreviations

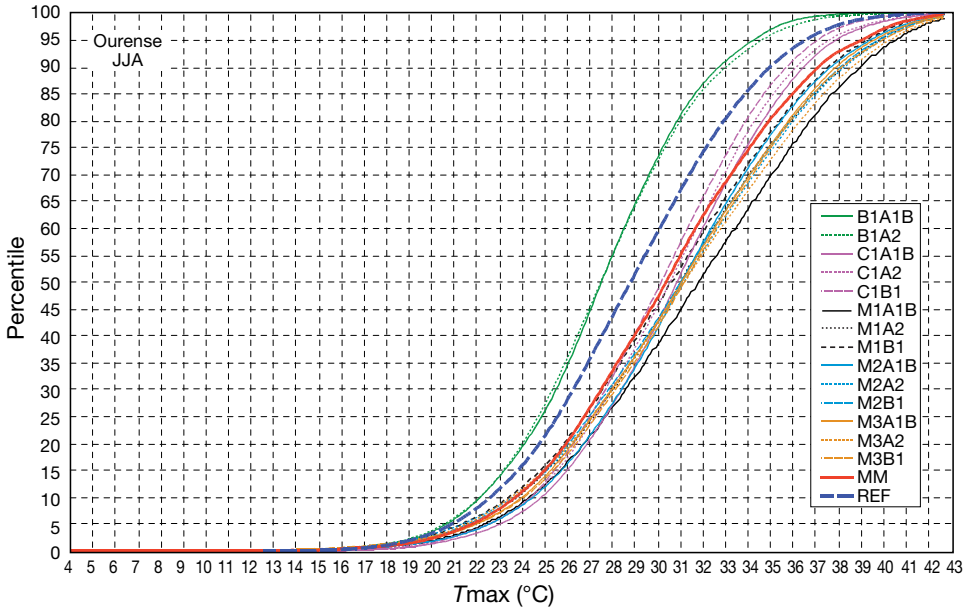


Fig. 7. Projections for daily maximum air temperatures (*T*max) in summer (JJA) at Ourense (reference period [REFREF]: JJA 1973–2002; scenario period [SCE]: JJA 2021–2050). MM: multi-model. See Table 2 for other abbreviations

seasons the MRW for *T*max was greater at continental (e.g. Oursense) than at maritime sites (e.g. A Coruña). This tendency also can be seen for *T*mean and, to a lesser degree, for *T*min.

Fig. 8c,d illustrates the frequency increase of daytime heat and Fig. 8e,f of night-time warm events, respectively. A value of 0% refers to a zero frequency increase. For the definitions of these rare events, see Table 3.

The less frequent the event in the REF, the higher its modelled frequency increase and the broader its uncertainty interval in the SCE. This tendency was

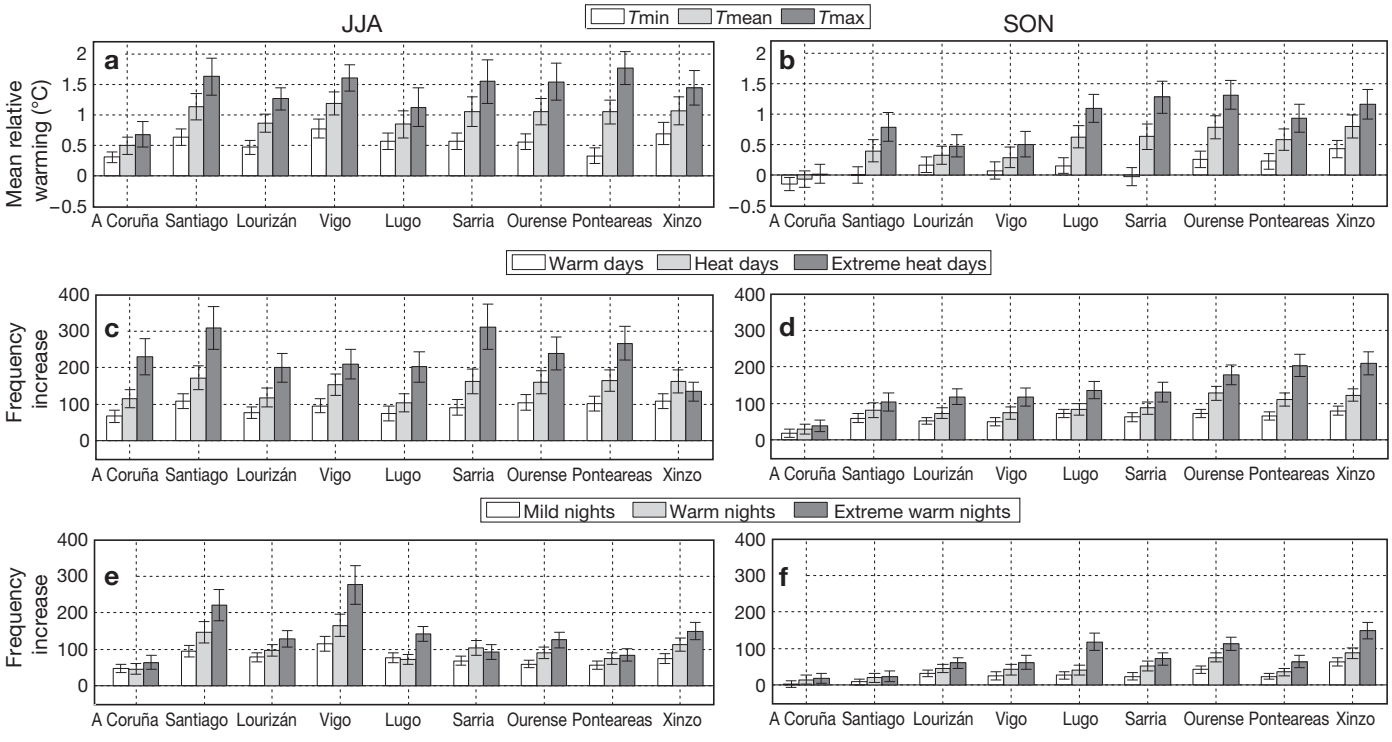


Fig. 8. (a,b) Mean relative warming (MRW) for daily minimum (*T*min), mean (*T*mean) and maximum (*T*max) air temperatures; (c,d) frequency increase of daytime heat events; and (e,f) frequency increase (% of reference period, REF) of night-time warm events at each of the 9 stations for (a,c,e) summer (JJA) and (b,d,f) autumn (SON). REF: JJA 1973–2002; scenario period (SCE): JJA 2021–2050. Bars in (a,b) refer to the MRW calculated from multi-models (MM) and errorbars to the corresponding standard errors. Bars and errorbars in (c,d,e,f) as in (a,b), but for frequency increases of warm events (see Table 3 for definitions)

more pronounced for daytime than for night-time heat/warm events (compare Fig. 8c,d to Fig. 8e,f), leading to a tripling to quadrupling of the modelled probability of extreme summer heat days accompanied by a large modelled uncertainty (Fig. 8c,d).

Finally, Fig. 9 shows the MM estimates of the frequency increase (*z*-axis) relative to the percentile thresholds used to define extreme events in the REF (*y*-axis, see Table 3) at the 9 stations under study (*x*-axis) in summer. Up to the 90th percentile, the relationship between the relative frequency of an extreme event in the REF and its frequency increase in the SCE was linear. For the highest percentiles this relationship changes to be exponential for daytime events (Fig. 9a) while it remains linear for their night-time correspondents at most of the stations (Fig. 9b).

## 5. DISCUSSION

It may be argued that the success in reproducing the observed percentiles is no proof of skill for the analogue method, as the observed distribution function would be equally met if the analogues had been selected randomly (Cubasch et al. 1996).

This is certainly true under optimal conditions where random number generators can be applied because the shape of the distribution is known. Nevertheless, under climate-change conditions, the predictand's distribution cannot be assumed *a priori* (Schär et al. 2004), and the unrestricted AN, used in the present study, has the remarkable advantage to simulate the CDF consistently with its GCM predictors. In addition, its accuracy can be compared with the results of Timbal et al. (2003), who used the restricted version of the AN.

On the other hand there are several problems yet to be solved.

Primarily, the artificial positive skewness of the summer CDFs in the scenario period, produced by the fact that the analogue search is restricted to the historical record (Imbert & Benestad 2005), could not be resolved in the present study. To handle this restriction, 2 criteria are suggested.

First, the maximum upper and lower percentiles that can be reliably reproduced under suboptimal conditions are to be identified. In our case we would state the upper threshold to be approximately at the 99th percentile for each of the 3 predictands (see MM in Figs. 3 to 5 and Figs. S4 & S5 in the supplement). This threshold should not be exceeded when defining a

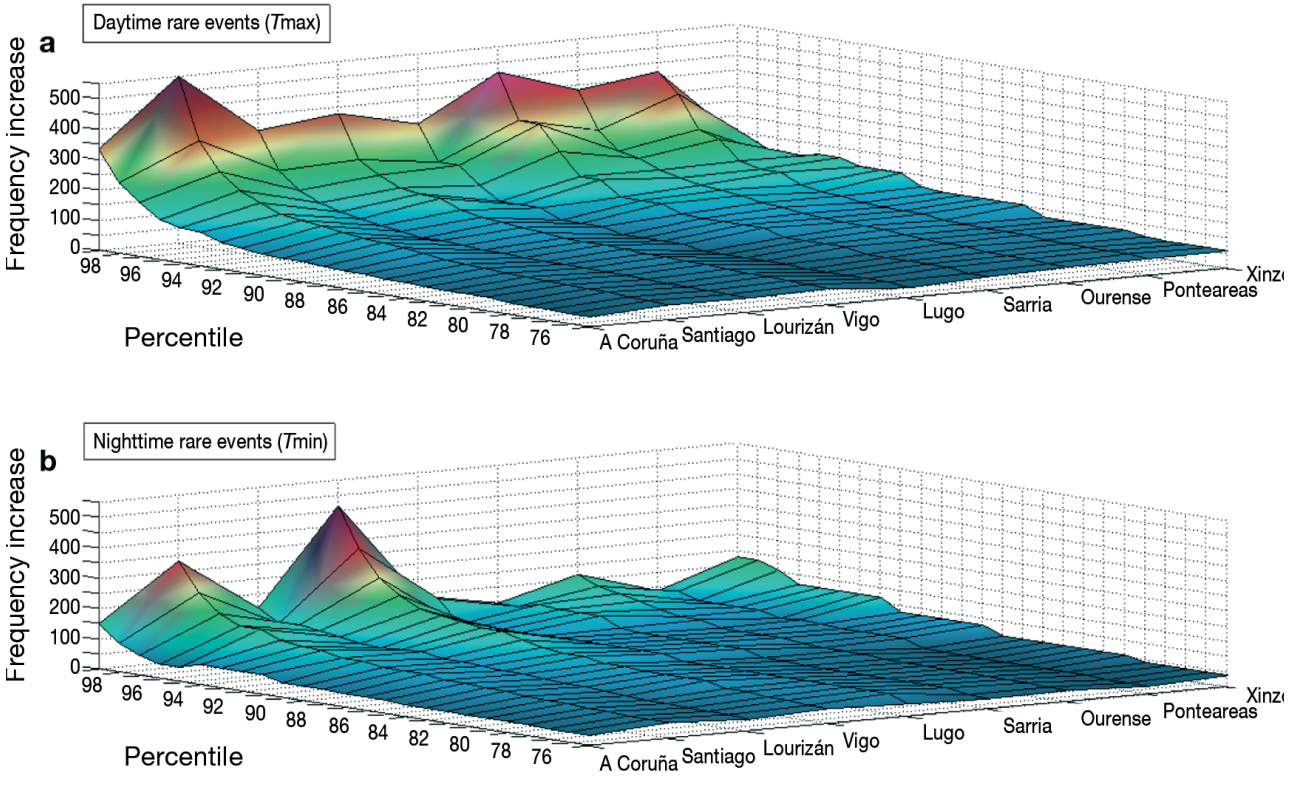


Fig. 9. Relationship between relative frequency in the reference period (REF) (*y*-axis) and frequency increase in scenario period (SCE) (*z*-axis) at all stations (*x*-axis) for (a) daytime heat events and (b) night-time warm events during summer (JJA). *T*max and *T*min: daily maximum and minimum air temperatures, respectively. Colours represent slope of surfaces

warm/heat event in the reference climate (see Table 3) and projecting its future frequency increase.

Second, the corresponding projections should be consistent with the expected increase of the first and/or second moments of the predictand's PDF (Schär et al. 2004), i.e. the percentiles should increase exponentially. In the present study, this criteria is approximately fulfilled up to the 99th percentile for *T*max (Fig. 9a) and for *T*mean (not shown). However, it is missed for *T*min which may either be a shortcoming of the AN, or result from increased simulated (clear sky) anticyclonic conditions, leading to a frequency increase of extreme daily temperature amplitudes.

Furthermore, more studies have to be undertaken to clearly separate the error of the SD technique under optimal conditions (Error 1) from the error introduced by the GCM control run (Error 2), both defining the combined error (Error 3). In the worst case, Error 1 and 2 compensate each other, yielding an artificial non-significant Error 3 (Huth 2004). In the present study, the good fit yielded in summer and autumn under suboptimal conditions may well have been affected by these compensation effects, which however were negligible, as Error 1 was acceptable.

One might further say that the good fit under suboptimal conditions is just a question of chance, since we chose a number of GCMs that yield reliable results in at least some seasons of the year. This is true to a certain extent, but it raises some interesting ideas. Using a larger multi-model ensemble and searching each member's optimal domain—which may differ from season to season (Wetterhall et al. 2005) and from predictor to predictor (Huth 2004)—it might be possible to reliably reproduce the distributions of the predictands in every season of the year. By optimising the domain the GCM predictors are extracted from, it may be geographically displaced from the best domain under optimal conditions. This is straightforward, as the regional centres of action simulated by GCMs rarely exactly coincide with the ones given by re-analysis data (Demuzere et al. 2009). It is therefore suggested to identify this displacement by applying classification schemes on both the re-analysis and GCM predictor data (Huth 2000) and to apply the AN on spatially differing domains in the training and validation period.

Finally, using the standard error to define the modelled uncertainty of future projections yields reasonable results. Although this choice is admittedly somewhat subjective, it should be seen as an easy-to-use option among several more sophisticated techniques (Katz 2002). Estimating the uncertainty interval of future climate projections is an issue of active scientific debate (Katz 2002, Stainforth et al. 2007, Knutti et al. 2010), and, as no clear consensus exists, every approach is necessarily subjective to a certain degree.

## 6. CONCLUSIONS

In the NW Iberian Peninsula, CDFs of local daily temperature variables can be reliably reconstructed by the AN in summer and autumn when an ensemble of control runs is downscaled and the resulting series are combined. The corresponding projections therefore neither have to be corrected nor given in relative terms to the control period. Their uncertainty interval is dominated by model errors, rather than by initial conditions or forcing uncertainties.

Besides the mean relative summer warming ranging from 0.7 ± 0.2°C to 1.8 ± 0.3°C for *T*max, a tripling to quadrupling of extreme heat days is projected. The latter is accompanied by a large modelled uncertainty.

For down-scaling GCM data, 2 important paradigms are highlighted:

(1) Harmonised daily GCM predictor data for a large multi-model ensemble should be made available for the downscaling community.

(2) Regional climate projections that ignore GCM uncertainty should be treated with caution, particularly if they refer to the probability of future extreme events. This holds especially if their corresponding control-run versions are not validated and/or the projections are given by the 'delta method', i.e. are presented relative to the control period.

*Acknowledgements.* We thank the World Data Center for Climate (WDCC) for providing the ENSEMBLES-stream 1 GCM data (www.mad.zmaw.de/wdc-for-climate/) and the European Centre for Medium-Range Weather Forecasts (ECMWF) for offering the ERA-40 re-analysis data (http://data-portal.ecmwf.int/data/d/era40_daily). Special thanks go to the Deutsche Klimarechenzentrum (DKRZ) (www.dkrz.de) staff, whose services made it possible to rapidly post-process the GCM data. The MeteoLab Toolbox has shown to be very useful for performing statistical downscaling techniques. We appreciate its free distribution (www.meteo.unican.es/en/software/meteolab) and we thank their authors at the Santander Meteorology Group. Last but not least, we thank 2 anonymous reviewers for their useful comments on an earlier version of the manuscript.

LITERATURE CITED

Barnett TP, Preisendorfer R (1978) Multifield analog prediction of short-term climate fluctuations using a climate state vector. J Atmos Sci 35:1771–1787

Benestad RE (2002) Empirically downscaled temperature scenarios for northern Europe based on a multi-model ensemble. Clim Res 21:105–125

Benestad RE (2005) Climate change scenarios for northern Europe from multi-model IPCC AR4 climate simulations. Geophys Res Lett 32:L17704 doi:10.1029/2005GL023401

Benestad RE, Hanssen-Baur I, Chen D (2008) Empirical-statistical downscaling. World Scientific Publishing, Singapore

Brandsma T, Buishand A (1998) Simulation of extreme precipitation in the Rhine basin by nearest neighbour resampling. Hydrol Earth Syst Sci 2:195–209

Bretherton CS, Smith C, Wallace JM (1992) An intercomparison

of methods for finding coupled patterns in climate data. J Clim 5:541–560

Christensen JH, Hewitson BC, Busuioc A, Chen A and others (2007) Regional climate projections. In: Solomon S, Qin D, Manning M, Chen Z and others (eds) Climate change 2007: the scientific basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 857–940

Cruz R, Lago A, Lage A, Rial ME, Díaz-Fierros F, Salson S (2009) Evolución recente do clima de Galicia—tendencias observadas en variables meteorolóxicas. In: Muñuzuri VP, Fernández Cañamero M, Gómez Gesteira JL (eds) Evidencias e impactos do cambio climático en Galicia. Consellería do Medio Ambiente, Santiago de Compostela, p 19–58

Cubasch U, Von Storch H, Waszkewitz J, Zorita E (1996) Estimates of climate change in Southern Europe derived from dynamical climate model output. Clim Res 7:129–149

Demuzere M, Werner M, Van Lipzig NPM, Roeckner E (2009) An analysis of present and future ECHAM5 pressure fields using a classification of circulation patterns. Int J Climatol 29:1796–1810

Drange H (2006a) ENSEMBLES BCCR-BCM2.0 20C3M run1, daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Entry.jsp?acronym=BCCR_BCM2.0_20C3M_1

Drange H (2006b) ENSEMBLES BCCR-BCM2.0 SRA1B run1, daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_BCM2_SRA1B_1_D

Drange H (2006c) ENSEMBLES BCCR-BCM2.0 SRA2 run1, daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_BCM2_SRA2_1_D

Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO (2000) Climate extremes: observations, modeling and impacts. Science 289:2068–2074

ECMWF (European Centre for Medium-Range Weather Forecasts) (2009a) ECMWF 40 year re-analysis (ERA-40) data archive. ECMWF, Reading. Available at: www.ecmwf.int/products/data/archive/descriptions/e4/index.html

ECMWF (European Centre for Medium-Range Weather Forecasts) (2009b) Meteorological parameters table 128: standard. ECMWF, Reading. Available at: www.ecmwf.int/services/archive/d/parameters

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall/CRC, Boca Raton, FL

Fowler HJ, Blenkinsop S, Tebaldi C (2007) Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Int J Climatol 27:1547–1578

Frías MD, Zorita E, Fernández J, Rodríguez-Puebla C (2006) Testing statistical downscaling methods in simulated climates. Geophys Res Lett 33:L19807 doi:10.1029/2006GL027453

Goodess C (2005) STARDEX, downscaling climate events. University of East Anglia, Norwich. Available at: www.cru.uea.ac.uk/projects/stardex/reports/STARDEX_FINAL_REPORT.pdf

Gutiérrez JM, Cofiño AS, Cano R, Rodríguez MA (2004) Clustering methods for statistical downscaling in short-range weather forecasts. Mon Weather Rev 132:2169–2183

Hewitson BC, Crane RG (1996) Climate downscaling: techniques and application. Clim Res 7:85–95

Huth R (1996) An intercomparison of computer-assisted circulation classification methods. Int J Climatol 16:893–922

Huth R (2000) A circulation classification scheme applicable in GCM studies. Theor Appl Climatol 67:1–18

Huth R (2004) Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors. J Clim 17:640–652

Imbert A, Benestad RE (2005) An improvement of analog model strategy for more reliable local climate change scenarios. Theor Appl Climatol 82:245–255

Jones PW (1999) First- and second-order conservative remapping schemes for grids in spherical coordinates. Mon Weather Rev 127:2204–2210

Karl TR, Wang WC, Schlesinger ME, Knight RW, Portman D (1990) A method of relating general circulation model simulated climate to observed local climate. I. seasonal statistics. J Clim 3:1053–1079

Katz RW (2002) Techniques for estimating uncertainty in climate change scenarios and impact studies. Clim Res 20:167–185

Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA (2010) Challenges in combining projections from multiple climate models. J Clim 23:2739–2758

Lall U, Sharma A (1996) A nearest-neighbour bootstrap for resampling hydrologic time series. Water Resour Res 32:679–693

Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130–141

Lorenz EN (1969) Atmospheric predictability as revealed by natural occurring analogues. J Atmos Sci 26:636–646

Lorenzo MN, Taboada J, Gimeno L (2008) Links between circulation weather types and teleconnection patterns and their influence on precipitation patterns in Galicia (NW Spain). Int J Climatol 28:1493–1505

Maraun D, Wetterhall F, Ireson AM, Chandler RE and others (2010) Precipitation downscaling under climate change. Recent developments to bridge the gap between dynamical models and the end user. Rev Geophys 48:RG3003 doi:10.1029/2009RG000314

Matulla C, Zhang X, Wang XL, Wang J, Zorita E, Wagner S, Von Storch H (2008) Influence of similarity measures on the performance of the analog method for downscaling daily precipitation. Clim Dyn 30:133–144

Nakicenovic N, Swart R (eds) (2000) Emission scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

Niehörster F, Fast I, Huebener H, Cubasch U (2008) The stream one ENSEMBLES projections of future climate change. ENSEMBLES Tech Rep 3. ENSEMBLES Project Office, Met Office, Exeter. Available at: http://ensembles-eu.metoffice.com/tech_reports.html

Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation, and confirmation of numerical models in the earth sciences. Science 263:641–646

PCMDI (Program for Climate Model Diagnosis and Intercomparison) (2009a) CMIP3 overview. Lawrence Livermore National Laboratory, Livermore, CA. Available at: http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php

PCMDI (Program for Climate Model Diagnosis and Intercomparison) (2009b) CMIP5 overview. Lawrence Livermore National Laboratory, Livermore, CA. Available at: http://cmip-pcmdi.llnl.gov/cmip5

Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20:4356–4376

Preisendorfer RW, Mobley CD (1988) Principal component analysis in meteorology and oceanography. Elsevier Academic Press, Amsterdam

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes in C++, 3rd edn. Cambridge University Press, Cambridge

Räisänen J (2007) How reliable are climate models? Tellus A 59:2–29

Randall DA, Wood RA, Bony S, Colman R and others (2007) Climate models and their evaluation. In: Solomon S, Qin D, Manning M, Chen Z and others (eds) Climate change 2007: the scientific basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 589–662

Roeckner E (2007a) ENSEMBLES ECHAM5-MPI-OM 20C3M run1, run2, run3 daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_20C3M_1_D, http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_20C3M_2_D, http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_20C3M_3_D

Roeckner E (2007b) ENSEMBLES ECHAM5-MPI-OM SRA1B run1, run2, run3 daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_SRA1B_1_D, http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_SRA1B_2_D, http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_SRA1B_3_D

Roeckner E (2007c) ENSEMBLES ECHAM5-MPI-OM SRA2 run1, run2, run3 daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_SRA2_1_D, http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_SRA2_2_D, http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_SRA2_3_D

Roeckner E (2007d) ENSEMBLES ECHAM5-MPI-OM SRB1 run1, run2, run3 daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_SRB1_1_D, http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_SRB1_2_D, http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_SRB1_3_D

Royer JF (2006a) ENSEMBLES CNRM-CM3 20C3M run1, daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_CNCM3_20C3M_1_D

Royer JF (2006b) ENSEMBLES CNRM-CM3 SRA1B run1, daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_CNCM3_SRA1B_1_D

Royer JF (2006c) ENSEMBLES CNRM-CM3 SRA2 run1, daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_CNCM3_SRA2_1_D

Royer JF (2006d) ENSEMBLES CNRM-CM3 SRB1 run1, daily values. CERA database. World Data Center for Climate, Hamburg. Available at: http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_CNCM3_SRB1_1_D

Schär C, Vidale PL, Lüthl D, Frei C, Häberli C, Liniger MA, Appenzeller C (2004) The role of increasing temperature variability in European summer heatwaves. Nature 427:332–336

Schmidli J, Goodess CM, Frei C, Haylock MR, Hundecha Y, Ribalaygua J, Schmith T (2007) Statistical and dynamical downscaling of precipitation: an evaluation and comparison of scenarios for the European Alps. J Geophys Res Atmos 112: D04105

Smith LA (2004) Predicting tomorrow's weather next week: the roles of uncertainty, probability, and model inadequacy in weather and in climate. AGU Fall Meeting, 13–17 Dec 2004, San Francisco, CA. Abstract NG22A-01. American Geophysical Union (AGU), Washington, DC. Available at: www.agu.org/cgi-bin/wais?bb=NG22A-01

Solomon S, Qin D, Manning M, Chen Z and others (eds) (2007) Climate change 2007: the scientific basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

Sordo C, Frías MD, Herrera S, Cofiño A, Gutiérrez JM (2008) Interval-based statistical validation of operational seasonal forecasts in Spain conditioned to El Niño-Southern Oscillation events. J Geophys Res Atmos 113:D17121

Stainforth DA, Allen MR, Tredger ER, Smith LA (2007) Confidence, uncertainty and decision-support relevance in climate predictions. Philos Trans R Soc Lond A 365:2145–2161

Teutschbein C, Wetterhall F, Seibert J (2011) Evaluation of different downscaling techniques for hydrological climate-change impact studies on the catchment scale. Clim Dyn (in press) doi:10.1002/joc.2256

Timbal B, Jones DA (2008) Future projections of winter rainfall in southeast Australia using a statistical downscaling technique. Clim Change 86:165–187

Timbal B, McAvaney BJ (2001) An analogue-based method to downscale surface air temperature application for Australia. Clim Dyn 17:947–963

Timbal B, Dufour A, McAvaney B (2003) An estimate of future climate change for western France using a statistical downscaling technique. Clim Dyn 20:807–823

Trigo RM, Palutikof JP (1999) Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach. Clim Res 13:45–59

Uppala SM, Kållberg PW, Simmons AJ, Andrae U and others (2005) The ERA-40 re-analysis. QJR Meteorol Soc 131: 2961–3012

Van den Dool H (1994) Searching for analogues, how long must we wait? Tellus A 46:314–324

Von Storch H (1999) On the use of 'inflation' in statistical downscaling. J Clim 12:3505–3506

Von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J Clim 6:1161–1171

Wetterhall F, Halldin S, Xu CY (2005) Statistical precipitation downscaling in central Sweden with the analogue method. J Hydrol 306:174–190

Wilby RL, Hay LE, Leavesley GH (1999) A comparision of downscaled and raw GCM output: implications for climate change in the San Juan River basin, Colorado. J Hydrol 225:67–91

Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods: supporting material of the Intergovernmental Panel on Climate Change. Task Group on Data and Scenario Support for Impacts and Climate Analysis, Rotherham

Wilks DS (2006) Statistical methods in the atmospheric sciences, 2nd edn. Elsevier Academic Press, Amsterdam

Zorita E, Von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J Clim 12:2474–2489

Zorita E, Hughes J, Lettenmaier D, Von Storch H (1995) Stochastic characterisation of regional circulation patterns for climate model diagnostics and estimation of local precipitation. J Clim 8:1023–1042

*Submitted: April 6, 2010; Accepted: July 7, 2010*

*Proofs received from author(s): July 8, 2011*