

On using principal components to represent stations in empirical–statistical downscaling

By RASMUSE. BENESTAD^{1*}, DELIANG CHEN², ABDELKADER MEZGHANI¹, LIJUN FAN³ and KAJSA PARDING¹, ¹*The Norwegian Meteorological Institute, PO Box 43, NO-0313 Oslo, Norway;* ²*University of Gothenburg, Gothenburg, Sweden;* ³*Key Laboratory of Regional Climate – Environment for East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*

(Manuscript received 23 April 2015; in final form 28 September 2015)

ABSTRACT

We test a strategy for downscaling seasonal mean temperature for many locations within a region, based on principal component analysis (PCA), and assess potential benefits of this strategy which include an enhancement of the signal-to-noise ratio, more efficient computations, and reduced sensitivity to the choice of predictor domain. These conditions are tested in some case studies for parts of Europe (northern and central) and northern China. Results show that the downscaled results were not highly sensitive to whether a PCA-basis or a more traditional strategy was used. However, the results based on a PCA were associated with marginally and systematically higher correlation scores as well as lower root-mean-squared errors. The results were also consistent with the notion that PCA emphasises the large-scale dependency in the station data and an enhancement of the signal-to-noise ratio. Furthermore, the computations were more efficient when the predictands were represented in terms of principal components.

Keywords: empirical–statistical downscaling, temperature, principal component analysis

To access the supplementary material to this article, please see Supplementary files under ‘Article Tools’.

1. Introduction

Some future planning and climate change adaptations require reliable information about future changes in local temperature or precipitation. Global climate models (GCMs) provide the best tools for making climate change projections (Solomon et al., 2007), but they are only designed to describe the large-scale features and do not provide sufficient details for many applications (Hellström et al., 2001). However, the local climate is linked to the large scales, and it is possible to make some inferences about local climate change through downscaling. Downscaling is often used to quantify the local response to global warming, either through the means of empirical–statistical downscaling (ESD) or regional climate models (RCMs) (Christensen et al., 2007). While there are various RCMs with different abilities to simulate local climate (Linden and Mitchell, 2009), ESD too may involve many different approaches and techniques

(Wilby et al., 1998a, 1998b; Easterling, 1999; Zorita and Storch, 1999; Huth, 2002; Salath, 2005; Benestad et al., 2008; Brands et al., 2013; Gutman et al., 2012). Indeed, the results of downscaling may vary depending on the chosen strategy or method, and hence the sensitivity of the results to different options must be examined (Hewitson et al., 2014). A relevant question for the projection of local climate change is how results from downscaling depend on assumptions and chosen strategies.

The optimal approach to deriving the local response to a large-scale climate change varies depending on the context. In some circumstances, the objective is not just to infer the local temperature at a single location, but one may be interested in the temperature over a region. A traditional approach has been to downscale one site at a time (Benestad, 2011; Fan et al., 2013), but it is also possible to represent the stations in terms of a principal component analysis (PCA) (Strang, 1988; Press et al., 1989; Wilks, 1995). The variations in the station records are expressed mathematically as a weighted combination of spatial patterns when PCA are used as a basis for downscaling, with different weights applied at different times (see Appendix). In this article,

*Corresponding author.
email: rasmus.benestad@met.no

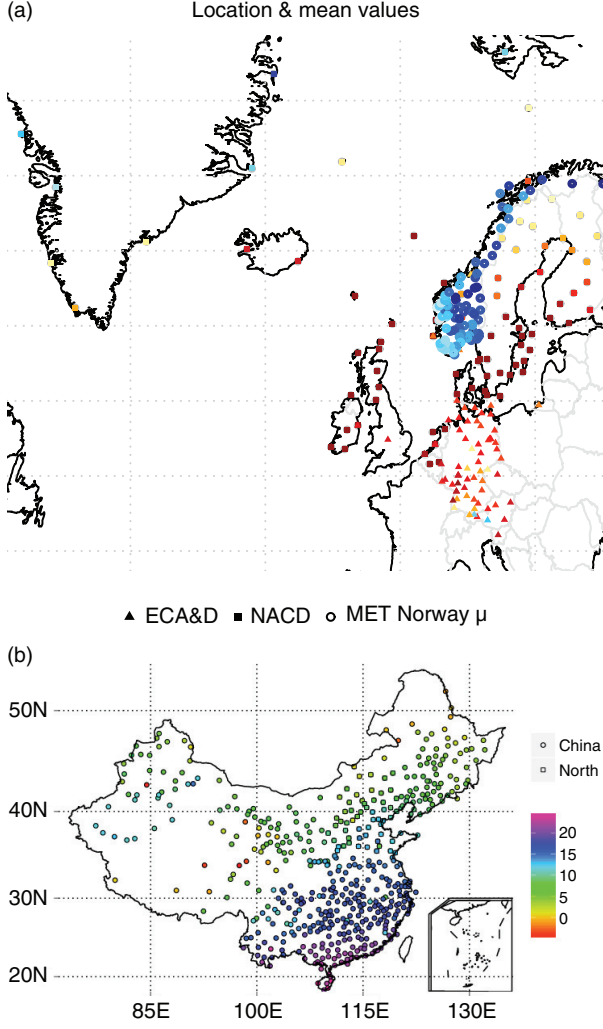


Fig. 1. The location of the thermometer readings and the rain gauge records (μ) over Europe (a) and the temperature stations from China (b). The colour coding is according to the mean values recorded.

we examine whether the traditional approach and the multi-site PCA-based approach to downscaling give comparable results, and if one outperforms the other.

1.1. Motivation for using the PCA-basis

Local manifestations of a global climate change imply a scale dependency, where the local temperature or precipitation ($y(\vec{r})$) can be expressed as a function of the large-scale region (X), geography ($g(\vec{r})$) and local processes (η):

$$y(\vec{r}) = f(X, g(\vec{r}), \eta),$$

where \vec{r} is a vector containing the coordinates.

It is possible to apply ESD techniques to a group of sites such as in methods based on canonical correlation analysis (CCA) or the analogue method (Barnett and Preisendorfer,

1987; Barnston, 1995; Fernandez and Saenz, 2003; Mo, 2003), for instance using PCA to represent the multi-sites data variability according to

$$Y = U\Lambda V^T,$$

where Y is a matrix containing a set of local time series $y(\vec{r})$, U is the left inverse that describes a set of spatial patterns describing the most pronounced spatially coherent variability (Benestad, 2013), Λ is a diagonal matrix holding the eigenvalues describing the prominence of each spatial pattern, and V^T is the transpose of the right inverse describing how strongly present each of the spatial patterns is at a given time. The left and right inverse satisfy $U^T U = V^T V = I$, where I is the identity matrix.

There are several potential benefits associated with PCA: (1) it could enhance the signal since it is designed to emphasise coherent spatial variations where the leading modes represent the patterns with maximum variance, (2) it accounts for the spatial covariance between the predictands at different locations, and (3) it can provide an effective means for downscaling several stations in a dense network if a reduced set of principal components (PCs) describes the large-scale dependent variability (higher order PCs may represent local noise η with no dependency to the large scales). The reduction in computational demand is valuable when the downscaling is applied to large ensembles of GCMs or RCMs, which is necessary in order to account for natural variations (Deser et al., 2012) and model differences (Allen et al., 2001; Meehl et al., 2005, 2007a, 2007b).

The application of a PCA-basis for downscaling calls for a research question: Are the downscaled results sensitive to the choice of using PCA rather than a more traditional approach? Hence, the hypothesis that we want to test is: $\mathcal{H}_0 = \text{PCA and traditional single-station downscaling provide comparable results}$. There are also additional hypotheses: $\mathcal{H}_1 = \text{PCA emphasises the large-scale dependency in the station data}$, and $\mathcal{H}_2 = \text{PCA makes the results less sensitive to the predictor domain}$. Hypothesis \mathcal{H}_1 implies that the higher order PCs represent local noise and can be addressed with the question of how many PCs are needed to recover most of the variability in observations. It is inspired by known mathematical properties of PCA (Strang, 1988; Press et al., 1989), where the leading modes represent the patterns associated with the highest variance that are usually spatially coherent. The higher order modes represent small-scale noise, and the PCA is expected to have a similar effect as filtering the data according to scales and variance.

Hypothesis \mathcal{H}_2 may be true as a consequence of \mathcal{H}_1 . The sensitivity of the results to the spatial domain size was examined through repeated downscaling using different domains. The assessment of the sensitivity was based on the

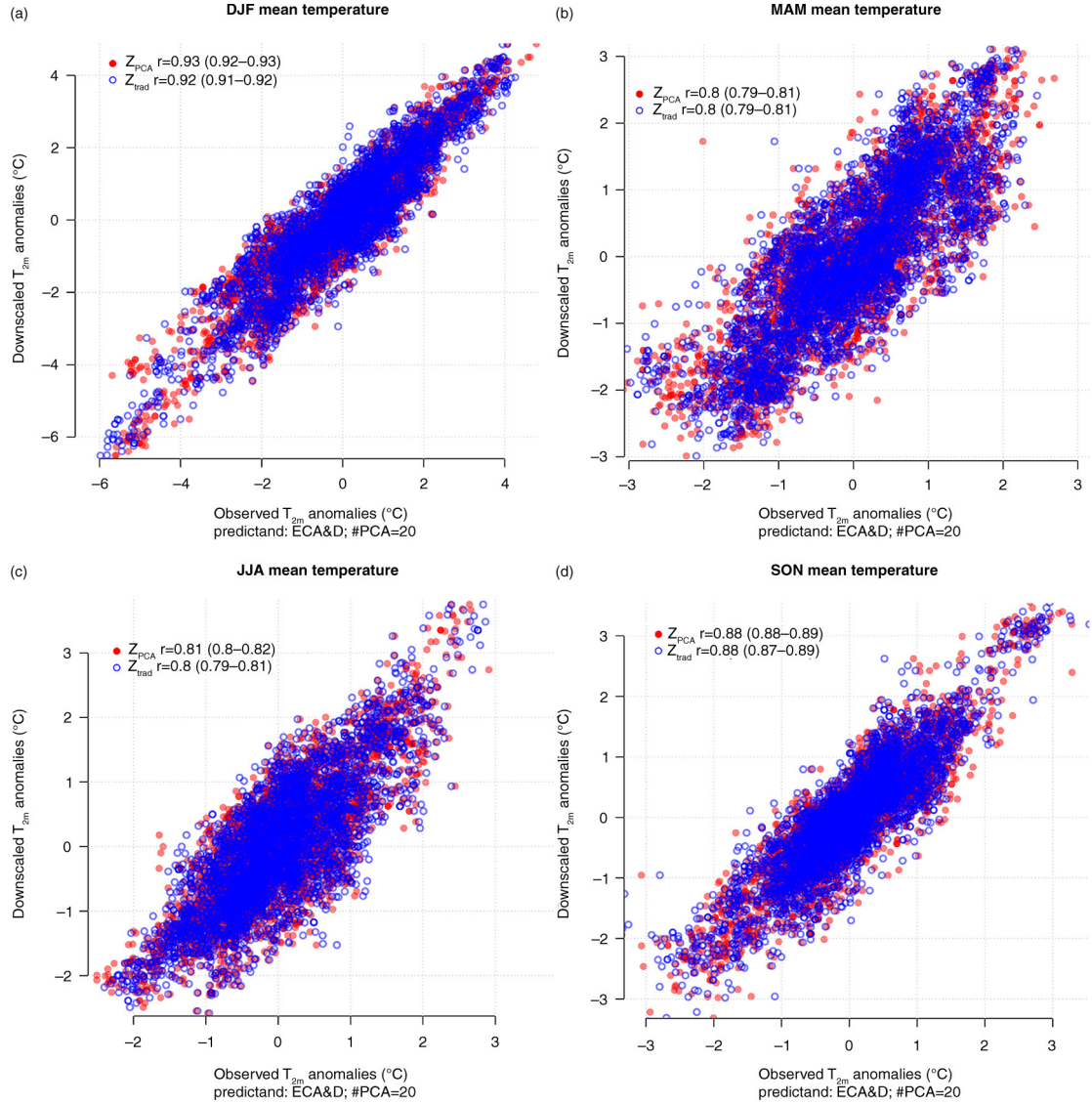


Fig. 2. Comparison between observed seasonal mean temperature anomaly and corresponding downscaled results using the PCA-based strategy and a traditional approach.

correlation between the downscaled and observed seasonal temperature anomalies for both downscaling strategies. Root-mean-squared errors (RMSE) were also computed for each station and were used to assess the sensitivity of two downscaling approaches to the choice of predictor domain size.

2. Data and method

Climate can, for all intents and purposes, be defined as the statistics of weather, and in any study of climate, a statistical sample of a weather variable should be sufficiently large to get a representative estimate of the probability distribution function (pdf) or a parameter describing some characteristics of the pdf. The present study analysed seasonal mean

temperature, as it was assumed that scales in time and space are connected. The statistical sample for the seasonal time scale is about 90 data points, although the real sample size is somewhat less due to autocorrelation. The mean estimates for seasons with less than 90 d of valid data were flagged as missing values.

We downscaled the winter, summer, spring and autumn mean temperature estimates independently from temperatures taken from either of two different European data sets in two experiments: the monthly mean North Atlantic Climate Dataset NACD (Frich et al., 1996) for the period 1890–1990 (not shown) and daily mean temperatures from the European Climate Assessment ECA&D (Klein Tank et al., 2002), respectively. Only the stations with complete

Table 1. Correlation score for downscaled seasonal mean temperature anomalies from the cross-validation

Area (km ²)	5 700 000	25 100 000	2 100 000	19 700 000	12 600 000	46 300 000
DJF						
PCA	0.94 (0.93, 0.94)	0.89 (0.89, 0.90)	0.93 (0.92, 0.93)	0.90 (0.90, 0.91)	0.93 (0.92, 0.93)	0.82 (0.81, 0.83)
trad.	0.92 (0.92, 0.93)	0.89 (0.88, 0.89)	0.92 (0.91, 0.92)	0.89 (0.88, 0.89)	0.91 (0.91, 0.92)	0.77 (0.76, 0.78)
JJA						
PCA	0.82 (0.80, 0.83)	0.77 (0.76, 0.79)	0.80 (0.79, 0.82)	0.78 (0.77, 0.80)	0.80 (0.79, 0.81)	0.74 (0.73, 0.76)
trad.	0.80 (0.79, 0.81)	0.75 (0.74, 0.77)	0.80 (0.79, 0.81)	0.76 (0.74, 0.77)	0.79 (0.77, 0.80)	0.69 (0.67, 0.71)
Domain						
W (°E)	−15.00	−50.00	−10.00	−70.00	−25.00	−90.00
E (°E)	25.00	40.00	20.00	30.00	35.00	50.00
S (°N)	45.00	35.00	50.00	40.00	40.00	30.00
N (°N)	65.00	75.00	60.00	67.00	70.00	75.00
N_{PCA}	20	20	20	20	20	20

The highest score for each trial is shown in bold and the numbers in the bracket indicate the 95% confidence interval for the correlation estimate. The standard deviation of the PCA-based scores was 0.07, whereas the standard deviation for the traditional method was 0.08, suggesting that the PCA-based ESD was less sensitive to the choice of predictor size.

data were retained. The predictor was taken from either of two different reanalyses: large-scale 2-metre air temperature from the ERA40 reanalysis (Simmons and Gibson, 2000; Bengtsson et al., 2004) and the NCEP/NCAR reanalysis I (Kalnay et al., 1996). The predictors were represented in terms of EOFs, and the seven leading modes (for PCA-based ESD, the number of EOFs representing the predictors is unrelated to the number of PCs representing the predictands). The results were sensitive to neither the choice of predictand data set nor the predictor data.

The experiment was repeated for China to check whether the results were robust with respect to region, where the winter, summer, spring and autumn mean temperatures were estimated from daily mean temperature. The observed daily mean temperature dataset from 825 stations in China was collected from the China Meteorological Data sharing service system (www.data.cma.gov.cn for details). Only 499 (whole of China) or 58 (northern China) stations with complete data over the period 1961–2012 were retained. The predictor was taken from large-scale monthly 2 m air temperature from NCEP/NCAR reanalysis.

To assess the robustness of the PCA-based ESD further, an additional set of downscaling was carried out for the annual wet-day mean μ (with threshold for wet-day at 1 mm/day) for 62 Norwegian rain gauge records taken from the Norwegian Meteorological Institute's climate archive (www.eklima.met.no). The set of station records for μ was complete for the period 1907–1997; however, since the NCEP/NCAR reanalysis only covered 1948–2014, the downscaling was carried out for 1948–1997 (50 yr). The saturation vapour pressure was used as predictor for μ , and was estimated using the Clausius–Clapeyron equation and the surface temperature taken from the reanalysis, as in Benestad and Mezghani (2015). The locations of all stations are shown in Fig. 1.

The downscaling used a stepwise multiple linear regression based on the `esd`-package (Benestad et al., 2014) for the R computing environment (R Core Team, 2014).¹ The downscaling analysis for the NACD data was carried out for the 1958–1990 period and 63 different stations in the Nordic countries as well as on the east coast of Greenland, the British Isles and in the Netherlands (not shown). The downscaling that involved the ECA&D data and NCEP/NCAR reanalysis are shown here, and the data spanned the years 1948–2013 and the region 11°W–21°E/46°N–64°N. The experiments were done with three different sets of PCAs (20, 10 and 4, respectively). Cross-validation was used to reduce the risk of artificial skill, whereby the data were split into five different equally sized sub-samples and where four sub-samples were used for calibration and the remaining was used for out-of-sample evaluation.

3. Results

A comparison between the results from traditional single-site downscaling applied to each of the stations separately and the PCA-based multi-site downscaling indicated similar results for all seasons. Figure 2 shows the downscaling results for ECA&D using the NCEP/NCAR reanalysis as predictor, but replacing either of these with the NACD or ERA40 gave the same results. The regression gave a good fit to all the stations and for the 20 PCs, and the results were quite robust to the choice of strategy.

The cross-validation correlation between the downscaled and observed seasonal mean anomaly temperatures was

¹The R-script with the source code is provided in the supporting material.

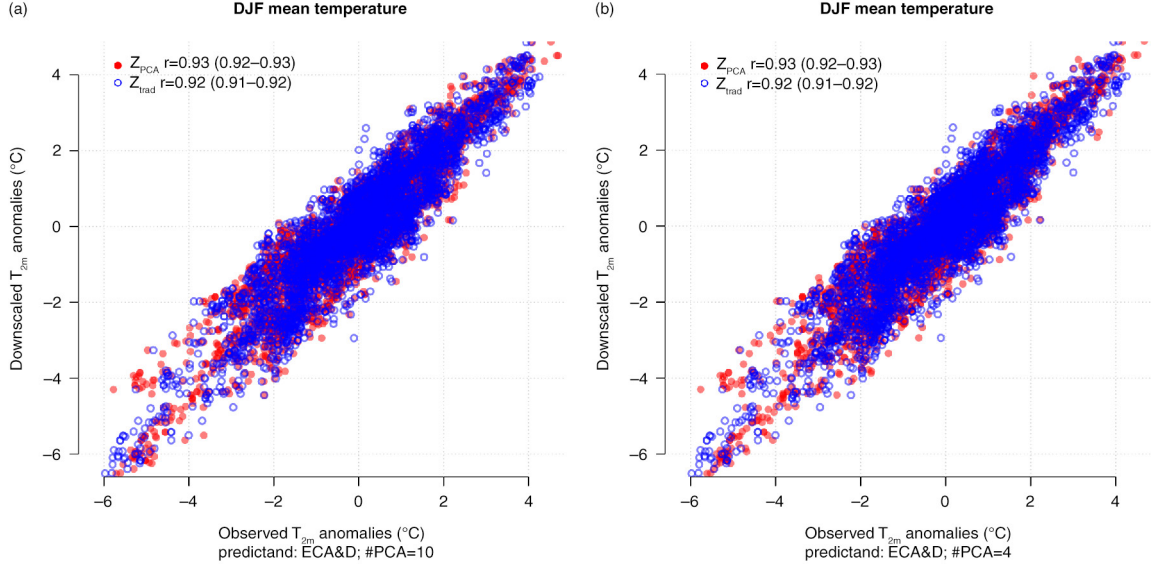


Fig. 3. Comparison between observed seasonal mean temperature anomaly and corresponding downscaled results using the PCA-based strategy and a traditional approach.

systematically higher for PCA-based downscaling than for traditional individual station downscaling. Where the correlation scores for the PCA-based results were 0.93, 0.80, 0.81 and 0.88 for winter, spring, summer and autumn, respectively, the traditional approach gave similar but marginally lower scores: 0.92, 0.80, 0.80 and 0.88. Table 1 shows similar scores for when the exercise shown in Fig. 2 was repeated with different predictor domain choices. The lowest cross-validation correlation score 0.69 and was associated with the largest predictor domain and the traditional downscaling approach. The highest correlation score was 0.94 for the PCA-based downscaling and the smallest predictor domain. The standard deviation of the cross-validation correlation scores for the set of exercises based on different predictor choices was 0.07 for the PCA-based strategy and 0.08 for the traditional approach, indicating that the downscaled results were more robust for the PCA-based approach (Table 1).

Figure 3 shows the downscaling repeated for smaller sets of PCs used to represent the station data. The correlation for all stations were comparable to the traditional single-site approach even when only four PCs were retained. In both these cases the cross-validation correlation was 0.93 as in the original exercise with 20 PCs (Fig. 2), and the higher order modes had little effect on these bulk scores for the entire set of stations.

However, some stations nevertheless showed greater discrepancy with observations for reduced PC sets. Figure 4 shows a more detailed comparison between the observations (black) and the results from the two downscaling strategies (PCA-based in red and traditional in blue) for a small selection of stations. All the panels except the upper right one show the downscaled temperature for the location with

the lowest correlation between observations and the PCA-based results. The upper panels show the results for PCA-based downscaling with only four EOFs, whereas the lower panels show results based on 10 and 20 EOFs, respectively. The upper right panel shows the results for PCA-based downscaling with four EOFs for the station with the highest correlation. Figures 3–4 all indicate similar results in general for the PCA-based and the traditional downscaling strategy, and the correlation for all stations and all years combined (shown in the figures) was 0.93 for the winter-time (DJF) PCA-based results and 0.92 for the corresponding traditional results (Fig. 2). The same statistics for the summer was 0.81 and 0.80.

The PCA-based downscaling preserved to a larger extent the spatial covariance of the seasonal mean anomalies of the station data than the downscaling for the individual stations. Figure 5 shows the covariance matrix for the observations (left) and the logarithm of the ratio of the corresponding downscaled results to the observations. The middle panel which shows the results for the PCA-based downscaling and exhibits smaller values for $\ln(\text{cov}(x_{DS})/\text{cov}(x_o))$ than the right panel displaying the results for the individual stations. For both cases, the covariance of downscaled results was lower than that of the observations (anomalies) because results from the regression model capture less than 100% of the variance. However, the correlation between the spatial structure in the covariance matrix was 0.99 for the PCA-based results and 0.98 for the traditional approach. Furthermore, the covariance matrix for the traditional downscaled results contained some elements with values greater than 1, also suggesting a slight degrading of the spatial covariance between the stations.

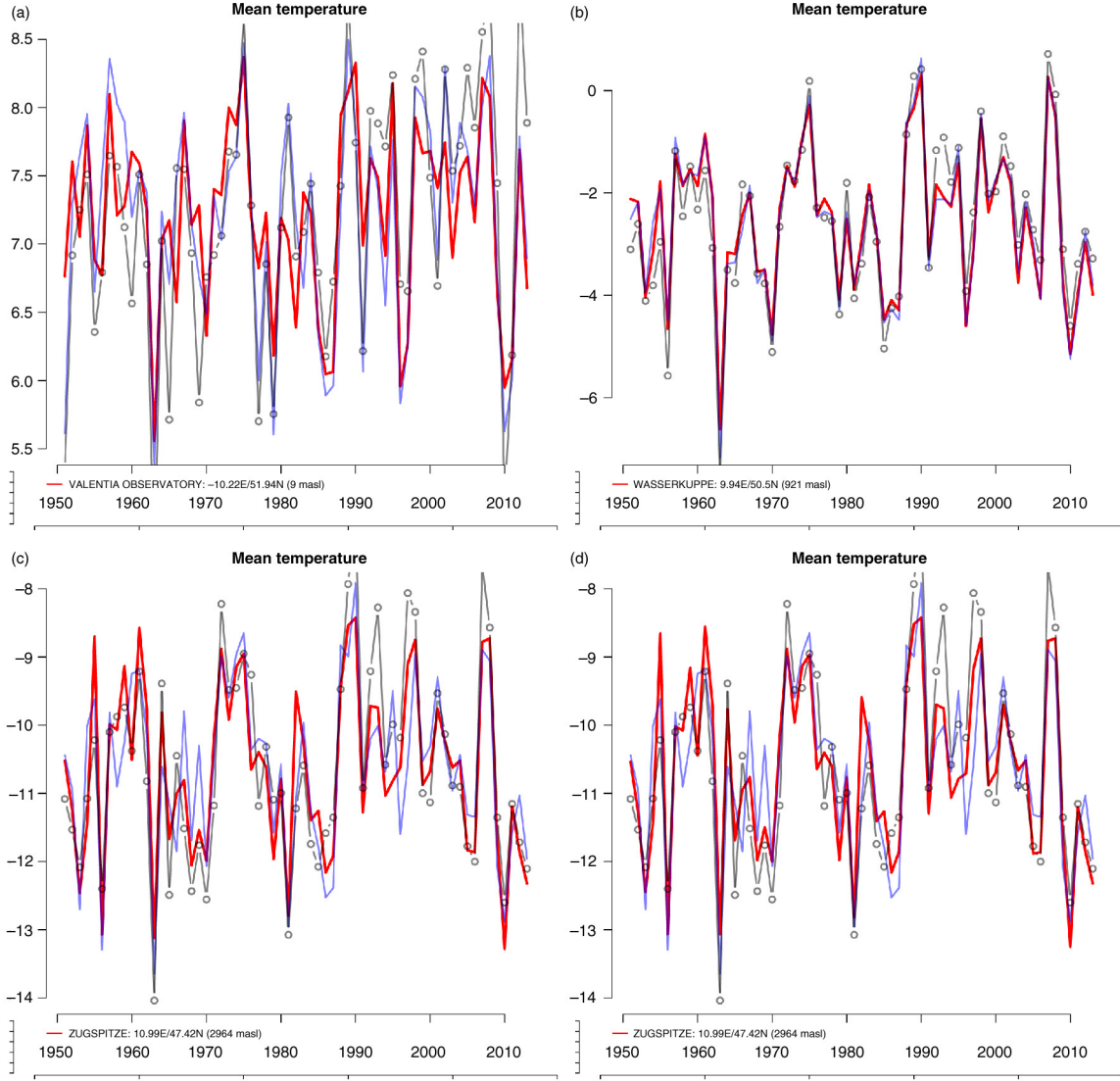


Fig. 4. Comparison between downscaled results and observations for a single station where (a) only 4 PCAs were retained and with the lowest correlation, (b) only 4 PCAs were retained and with the highest correlation, (c) only 10 PCAs were retained and with the lowest correlation, and (d) only 20 PCAs were retained and with the lowest correlation. The observed data are shown in black, the PCA-based downscaled results in red, and the traditional downscaled results in blue.

To further test \mathcal{H}_1 , we repeated the PCA-based downscaling exercise after excluding the four leading PCA modes (Fig. 6). The multi-station-multi-year correlation for the PCA-based results was reduced to 0.08 in winter and 0.12 in summer. Hence, most of the signal connected to the large scales was embedded in the four leading PCs.

Another measure of skill that is sensitive to mean biases and the magnitudes of the downscaled results is the RMSE, where high skill is associated with low RMSE. Figure 7 shows a comparison between the RMSE estimated station-wise for the PCA-based and traditional approach, plotted against each other. The results indicate higher RMSE for the traditional approach as the points in the scatter plot tend to lie above

the diagonal (Fig. 7). The RMSE scores were taken from the same tests as presented in Table 1 for a number of different domain sizes.

Using PCA to represent predictands instead of individual station records reduces the downscaling time. It took 18 s to downscale 62 stations over 65 yr and one season using the PCA-basis with 20 PCs (Fig. 2; on a laptop with a 64-bit Intel Core i7-3540M CPU running at 3.00 GHz \times 4 and 15.5 GiB of RAM), compared to 30 s for the more traditional approach downscaling each station individually.

Tables 2 and 3 show the analysis repeated for China and northern China, respectively, and suggest a similar tendency

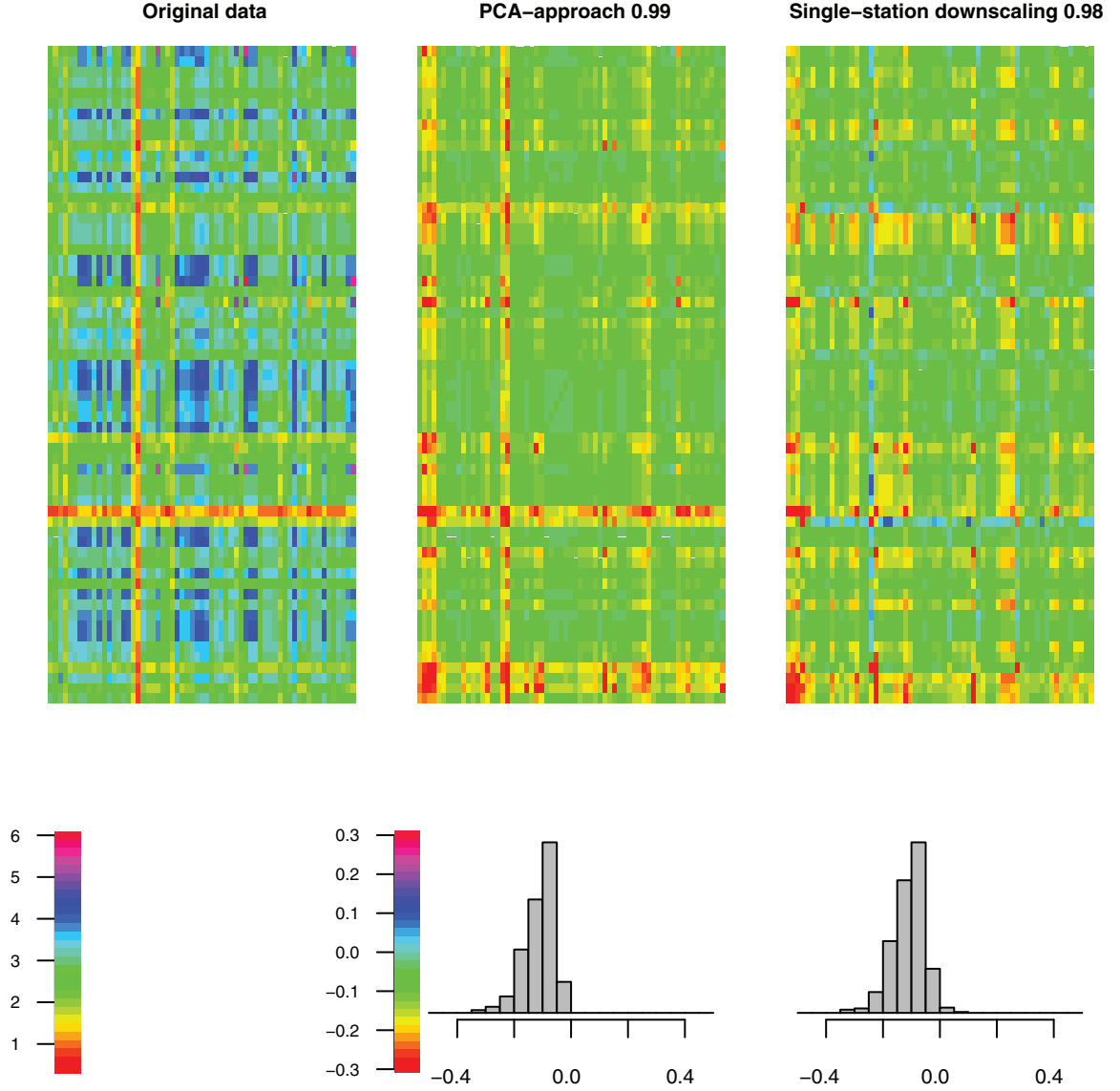


Fig. 5. The covariance of the seasonal mean anomalies of the station data x_o (left) and the difference between the two downscaling strategies in terms of $\ln(\text{cov}(x_{DS})/\text{cov}(x_o))$ (middle and right).

for the PCA-based ESD to be associated with slightly higher scores as in northern Europe. Hence, the main finding that the PCA-based strategy was at least as good as the traditional method was not specific for the European region. A similar experiment carried out for the annual wet-day mean μ for 62 locations in Norway gave a more even performance between the PCA-based and traditional approach (Table 4). In this set of experiments, the trials with the same predictor region were repeated three times but with different number of PCAs retained for the stations and, hence, were not independent in terms of predictor domains (the downscaled results for the traditional approach was identical in all these three trials). When taking into account the confidence estimates in Table 1, there were six trials (50%) where

the PCA-based approach gave higher scores and where the respective 95% confidence intervals did not overlap and none where the traditional approach was associated with significantly higher skill than the PCA-based approach. The corresponding assessment for μ in Table 4 indicated two cases where the PCA-based approach gave significantly higher scores and where the respective 95% confidence intervals did not overlap and none where the traditional approach gave significantly higher correlation.

4. Discussion

The downscaled results were not very sensitive to the use of PCA, and hypothesis \mathcal{H}_0 could therefore not be rejected.

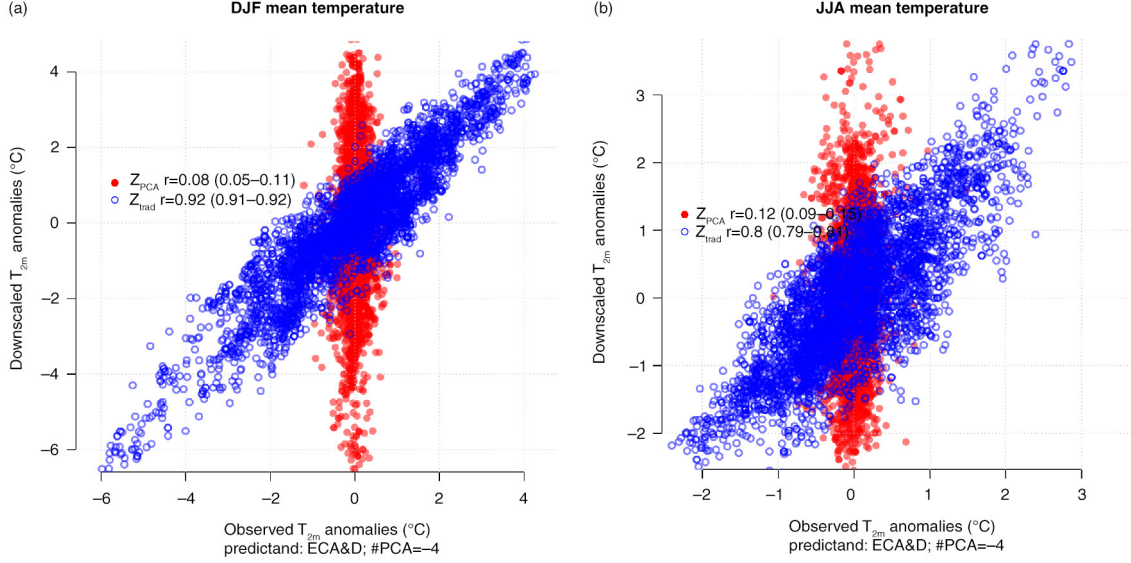


Fig. 6. Comparison between observed seasonal mean temperature anomaly and corresponding downscaled results using the PCA-based strategy and a traditional approach after having excluded the four leading PCA modes. Left panel shows results for the winter and right for the summer.

However, the PCA-based skill scores were marginally higher than the results from downscaling individual stations in all the tests. Performing downscaling with different numbers of PCs retained for the station PCA also suggested that most of the dependency to the large scales resided in the leading modes. Hence \mathcal{H}_1 also was consistent with these results. Finally, the results from the exercise with repeated

analysis based on different predictor domains indicated less sensitivity for the PCA-based approach, in agreement with hypothesis \mathcal{H}_2 .

The results presented in this study suggest that the leading PCs actually capture the large-scale dependency of local temperature variability in northern Europe. Downscaling using PCA to represent the predictand offers a way of making

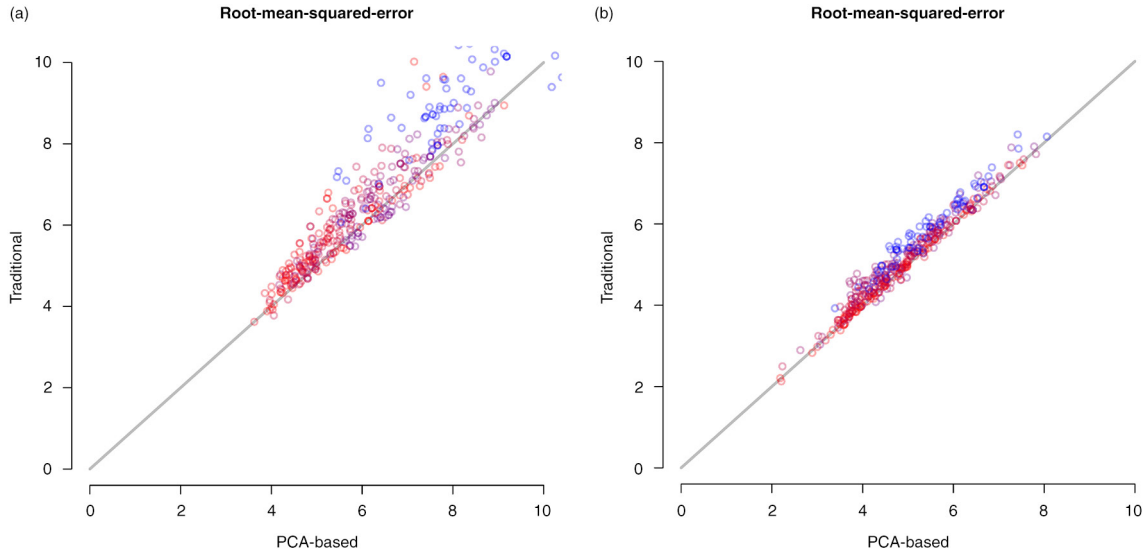


Fig. 7. Comparison between station-wise RMSE for a number of downscaling exercises with different domain sizes. The PCA-based results are plotted along the x-axis and results from the traditional approach along the y-axis. The units are degree C (anomalies). Points above the diagonal indicate higher skill for the PCA-based downscaling strategy. The different colours represent different predictor domain. Left panel shows results for the winter and right for the summer.

Table 2. Skill scores for the PCA-based and traditional ESD for China

Area (km ²)	27 800 000	31 800 000	35 800 000	34 400 000	39 300 000	44 200 000
DJF						
PCA	0.87	0.86	0.87	0.87	0.86	0.86
trad.	0.83	0.82	0.82	0.84	0.83	0.82
JJA						
PCA	0.71	0.69	0.72	0.73	0.70	0.70
trad.	0.70	0.71	0.70	0.68	0.65	0.64
N_{PCA}	20	20	20	20	20	20
Domain						
W (°E)	70.00	65.00	60.00	70.00	65.00	60.00
E (°E)	140.00	145.00	150.00	140.00	145.00	150.00
S (°N)	15.00	15.00	15.00	10.00	10.00	10.00
N (°N)	55.00	55.00	55.00	60.00	60.00	60.00
N_{PCA}	20	20	20	20	20	20

The entries are correlation scores from the cross-validation. The predictor was the NCEP/NCAR reanalysis and predictands included 499 stations with complete time series over the period 1961–2012. The highest score in each of the comparisons is shown in bold.

more efficient calculations and accounting for the spatial covariance. Higher order PCs can describe noise and, in conjunction with weather generators, could be used to simulate stationary stochastic small-scale variability through a Monte-Carlo approach. Furthermore, the leading modes of PCA can be used as a diagnostic for large-scale dependency. In other words, using PCA to represent the predictands can be useful when the signal is weak because it will emphasise the large scales in the leading modes and hence separate the signal from noise.

The number of PCs needed to reproduce the observations is expected to vary with the chosen element, season, the extent of the area represented by the stations, and the region. This depends on the proportion of the total variance accounted for by the leading modes. In general, only a small number of PCs is needed for predictands representing spatially coherent variations as, for example,

the winter temperature in northern Europe which was sufficiently represented by four PCs. In other situations, such as for the less spatially coherent precipitation, a small number of PCs may be insufficient.

The PCA-basis is not expected to be appropriate if the stations are too widely spread and involve different regional climate phenomena.

5. Conclusions

In summary, the results from a series of experiments suggested only marginally yet systematically higher correlation between observed seasonal mean temperature anomalies and corresponding downscaled results for PCA-based ESD. This finding was true for both cross-validated results as well as downscaled result based on the calibration data. Our results were consistent with the notion that PCA emphasises the

Table 3. The results of the PCA-based ESD test for Northern China

Area (km ²)	8 400 000	3 800 000	14 900 000	2.3×10^7	23 900 000	5 700 000
DJF						
PCA	0.89	0.89	0.89	0.88	0.88	0.90
trad.	0.88	0.89	0.86	0.86	0.82	0.89
JJA						
PCA	0.81	0.82	0.79	0.74	0.75	0.79
trad.	0.74	0.80	0.72	0.68	0.69	0.78
Domain						
W (°E)	100.00	105.00	95.00	90.00	80.00	100.00
E (°E)	130.00	125.00	135.00	140.00	140.00	130.00
S (°N)	25.00	30.00	20.00	15.00	15.00	30.00
N (°N)	55.00	50.00	60.00	65.00	55.00	50.00
N_{PCA}	20	20	20	20	20	20

The predictor was the NCEP/NCAR reanalysis and predictands included 58 stations for North China. Numbers in bold face mark the highest score in each test.

Table 4. The results for PCA-based and traditional ESD for the annual wet-day mean precipitation taken from 62 Norwegian rain gauge records over the period 1907–1997

Area (km ²)	5 700 000	5 700 000	5 700 000	25 100 000	2 100 000	19 700 000	12 600 000	46 300 000
PCA	0.23 (0.2, 0.26)	0.23 (0.2, 0.27)	0.24 (0.21, 0.27)	0.26 (0.22, 0.29)	0.2 (0.17, 0.24)	0.31 (0.28, 0.34)	0.18 (0.14, 0.21)	0.33 (0.29, 0.36)
trad.	0.26 (0.23, 0.3)	0.26 (0.23, 0.3)	0.26 (0.23, 0.3)	0.15 (0.12, 0.18)	0.24 (0.2, 0.27)	0.27 (0.23, 0.3)	0.19 (0.16, 0.22)	0.26 (0.22, 0.29)
Domain								
W (°E)	−15	−15	−15	−50	−10	−70	−25	−90
E (°E)	25	25	25	40	20	30	35	50
S (°N)	45	45	45	35	50	−70	−25	−90
N (°N)	65	65	65	75	60	30	35	50
N_{PCA}	20	4	10	20	20	20	20	20

The entries are correlation scores from the cross-validation. Note that the only difference for the three first entries was retaining different numbers of PCAs for the stations (same predictor region). The predictor was the saturation water pressure estimated from the NCEP/NCAR reanalysis 2-metre temperature based on the Clausius–Clapeyron equation. The highest value for each test is shown in bold.

large-scale dependency in the station data and therefore gives a better reproduction of the original data. We also found the PCA-based approach to be more efficient and less sensitive to the choice of predictor domain than corresponding analysis for each station individually. Our conclusion is therefore that ESD can benefit from the use of PCA as a means of accounting for the predictands.

6. Acknowledgement

This work has been supported by COST-VALUE STSM, the Norwegian Meteorological Institute, CHASE-PL (EEA grant PL12-0104), SPECS (EU Grant Agreement 3038378), Swedish MERGE, BECC and VR. It was also meant to contribute towards efforts within CORDEX-ESD experiments.

7. Appendix

Regression

The regression for the PCA-strategy was similar to Benestad (2013):

$$\hat{y}_j = \beta_0 + \sum_{i=1}^n \beta_i x_i. \quad (\text{A1})$$

The regression was applied to each principal component in the same fashion as it was applied to each station record for the traditional downscaling strategy. The main difference was that the original temperatures were recovered from the singular value decomposition (Strang, 1988; Press et al., 1989) used to estimate the EOFs:

$$X = U\Lambda V^T, \quad (\text{A2})$$

where X represents the data matrix holding the original seasonal mean temperature, U holds the EOFs, Λ the eigenvalues, and V the principal components (PCs). The difference between the PCA-based and traditional strategies was that in the former, y_j in eq. (A1) was taken to be the respective PCs

[columns of V in eq. (A2)] whereas in the latter was the original data [columns of X in eq. (A2)].

The area of a sector was taken as $A = \int \text{acos}(\Phi) ad\Phi \times \Delta\theta / (2 * \pi)$, where a is the mean radius of the Earth (6378 km), θ is the longitude of the western and eastern boundaries in radians and Φ is latitude of the southern and northern boundaries. This integral can be solved analytically, and the area was computed according to:

$$A = a^2 [\sin(\Phi_{\max}) - \sin(\Phi_{\min})] \Delta\theta. \quad (\text{A3})$$

References

- Allen, M., Raper, S. and Mitchell, J. 2001. Uncertainty in the IPCC's third assessment report. *Science*. **293**, 430–433.
- Barnett, T. P. and Preisendorfer, R. W. 1987. Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Weather Rev.* **115**, 1825–1850.
- Barnston, A. 1995. Brief summary of NMC's canonical correlation analysis (CCA), optimal climate normals (OCN), and NMC coupled model forecasts for U.S. surface climate. *NOAA Exp. Long-Lead Forecast Bull.* **4**, 45–46.
- Benestad, R. and Mezghani, A. 2015. On downscaling probabilities for heavy 24-hr precipitation events at seasonal-to-decadal scales. *Tellus A*. **67**, 25954. DOI: <http://dx.doi.org/10.3402/tellusa.v67.25954>
- Benestad, R. E. 2011. A new global set of downscaled temperature scenarios. *J. Clim.* **24**(8), 2080–2098.
- Benestad, R. E. 2013. Association between trends in daily rainfall percentiles and the global mean temperature. *J. Geophys. Res. Atmos.* **118**(19), 10802–10810.
- Benestad, R. E., Hanssen-Bauer, I. and Chen, D. 2008. *Empirical–Statistical Downscaling*. World Scientific, Singapore.
- Benestad, R. E., Mezghani, A. and Parding, K. M. 2014. esd – the empirical–statistical downscaling tool & its visualization capabilities. MET report 11/15. figshare. Online at: <http://www.github.com/metno/esd> & DOI: <http://dx.doi.org/10.6084/m9.figshare.1160493>
- Bengtsson, L., Hodges, K. I. and Hagemann, S. 2004. Sensitivity of the ERA40 reanalysis to the observing system: determination

- of the global atmospheric circulation from reduced observations. *Tellus A.* **56**, 456–471.
- Brands, S., Herrera, S., Hernández, J. and Gutiérrez, J. M. 2013. How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa? A performance comparison for the downscaling community. *Clim. Dynam.* **41**, 803–817.
- Christensen, J. H., Hewitson, B., Busuioc, A., Chen, A., Gao, X. and co-authors. 2007. *Climate Change: The Physical Science Basis*. Cambridge University Press, New York.
- Deser, C., Knutti, R., Solomon, S. and Phillips, A. S. 2012. Communication of the role of natural variability in future North American climate. *Nat. Clim. Change*. **2**, 775–779. DOI: <http://dx.doi.org/10.1038/nclimate1562>
- Easterling, D. R. 1999. Development of regional climate scenarios using a downscaling approach. *Clim. Change*. **41**, 615–634.
- Fan, L., Chen, D., Fu, C. and Yan, Z. 2013. Statistical downscaling of summer temperature extremes in Northern China. *Adv Atmos. Sci.* **30**, 1085–1095.
- Fernandez, J. and Saenz, J. 2003. Improved field reconstruction with the analog method: searching the CCA space. *Clim. Res.* **24**, 199–213.
- Frich, P., Alexandersson, H., Ashcroft, J., Dahlström, B., Demare, G. R. and co-authors. 1996. *North Atlantic Climatological Dataset (NACD Version 1) – Final Report*. Scientific report 1. DMI, Copenhagen, Denmark.
- Gutman, E. D., Rasmussen, R. M., Liu, C., Ikeda, K., Gochis, D. J. and co-authors. 2012. A comparison of statistical and dynamical downscaling of winter precipitation over complex terrain. *J. Clim.* **25**, 262–281.
- Hellström, C., Chen, D., Achberger, C. and Räisänen, J. 2001. Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. *Clim. Res.* **19**, 45–55.
- Hewitson, B. C., Daron, J., Crane, R. G., Zermoglio, M. F. and Jack, C. 2014. Interrogating empirical–statistical downscaling. *Clim. Change*. **122**(4), 539–554.
- Huth, R. 2002. Statistical downscaling of daily temperature in central Europe. *J. Clim.* **15**, 1731–1742.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D. and co-authors. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471.
- Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G. and co-authors. 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol.* **22**, 1441–1453.
- Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B. and co-authors. 2007a. The WCRP CMIP3 Multimodel dataset: a new era in climate change research. *Bull. Am. Meteorol. Soc.* **88**, 1383–1394.
- Meehl, G. A., Covey, C., McAvaney, B., Latif, M. and Stouffer, R. J. 2005. Overview of the coupled model intercomparison project. *Bull. Am. Meteorol. Soc.* **86**, 89–93. DOI: <http://dx.doi.org/10.1175/BAMS-86-1-89>
- Meehl, G. A., Stocker, T. F., Idlingstein, W. D., Gaye, A. T., Gregory, J. M. and co-authors. 2007b. *Climate Change: The Physical Science Basis*. Cambridge University Press, New York.
- Mo, K. C. 2003. Ensemble canonical correlation prediction of surface temperature over the United States. *J. Clim.* **16**, 1605–1683.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. 1989. *Numerical Recipes in Pascal*. Cambridge University Press, Cambridge, UK.
- R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Salath, E. P. 2005. Downscaling simulations of future global climate with application to hydrologic modelling. *Int. J. Climatol.* **25**, 419–436.
- Simmons, A. J. and Gibson, J. K. 2000. *The ERA-40 Project Plan*. ERA-40 Project Report Series 1. ECMWF. Online at: <http://www.ecmwf.int>
- Solomon, S., Quin, D., Manning, M., Chen, Z., Marquis, M. and co-authors. 2007. *Climate Change: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, New York.
- Strang, G. 1988. *Linear Algebra and its Application*. Harcourt Brace & Company, San Diego, CA.
- van der Linden, P. and Mitchell, J. F. B. (eds.). 2009. *Ensembles: Climate Change and its Impacts: Summary of Research and Results from the ENSEMBLES Project*. Met Office Hadley Centre, Exeter.
- Wilby, R. L., Hassan, H. and Hanaki, K. 1998a. Statistical downscaling of hydrometeorological variables using general circulation model output. *J. Hydrol.* **205**, 1–19.
- Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C. and co-authors. 1998b. Statistical downscaling of general circulation model output: a comparison of methods. *Water Resour. Res.* **34**, 2995–3008.
- Wilks, D. S. 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, Orlando, FL.
- Zorita, E. and von Storch, H. 1999. The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *J. Clim.* **12**, 2474–2489.