INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **27**: 649–665 (2007)
Published online 9 October 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.1421

RMetS
ROYAL METEOROLOGICAL SOCIETY

An evaluation of statistical models for downscaling precipitation and their ability to capture long-term trends

R. E. Benestad,* I. Hanssen-Bauer and E. J. Førland
The Norwegian Meteorological Institute, 0313, Oslo, Norway

Abstract:

Large-scale changes in the sea-level pressure do not necessary reflect changes in the atmospheric moisture budget, and hence may not give a good representation of changes in precipitation as a result of a global warming. Statistical models that use both sea-level pressure and large-scale precipitation as predictors are evaluated for a number of locations in Fennoscandia. The statistical models in most cases were capable of capturing 60–80% of the year-to-year seasonal variations in precipitation, and a correlation analysis over independent data indicated predictive correlation scores in the range 0.2–0.5. A comparison between statistical models based on large-scale precipitation, sea-level pressure, and a mixture of these, indicated similar skills in terms of variance and predictive skill of inter-annual variations. Analyses of their ability to capture recent precipitation trends reveal potential problems regarding reconstructing long-term changes in the past. One explanation for the statistical models not giving similar past trend values as given by the station observations may be partly because the precipitation trends during the most recent 50 years are not well defined since the interval is not sufficiently long. This is supported by the fact that trend analysis for station observations based on two different data products, and different trend analysis strategies, do not correspond well with each other. An analysis for possible non-stationarities between large and local spatial scales does not indicate any significant presence of non-stationarities. Copyright © 2006 Royal Meteorological Society

KEY WORDS empirical-statistical downscaling; precipitation trend evaluation

Received 4 May 2005; Revised 21 July 2006; Accepted 26 July 2006

INTRODUCTION

An increase in global surface air temperature is arguably the most robust climate signal resulting from rising atmospheric concentrations of greenhouse gases (Houghton *et al*., 2001); however, changes in precipitation may lead to more serious impacts on human activities and biota in many regions. Furthermore, local rainfall is often more important than the mean precipitation over a greater region such as a global climate model (GCM) grid box. Therefore, it is important to infer local precipitation through the means of so-called 'downscaling' in GCM studies. Even in regional climate models (RCM), the spatial resolution is too coarse for many purposes, especially in complex terrain where precipitation may vary considerably over a few kilometres. In such areas where long-term climate observational series exist, IPCC (1995) recommended the application of empirical-statistical downscaling (hereby called 'empirical downscaling') to scenarios from GCMs to produce local description with fine-scale structures. Hence, development of methods for downscaling precipitation scenarios from GCMs is important.

A variety of empirical downscaling studies were reported by the Intergovernmental Panel on Climate Change (IPCC) (2001), and several studies including empirical downscaling of precipitation have been performed for Fennoscandia during the later years (Busuioc *et al*., 2001; Hellström *et al*., 2001; Benestad, 2002; Hanssen-Bauer *et al*., 2003; Hellström, 2003; Linderson *et al*., 2004; ; Chen *et al*., 2005; Imbert and Benestad, 2005). Comparisons of results from different downscaling techniques (Hellström *et al*., 2001; Hanssen-Bauer *et al*., 2003), different choices of domain size (Benestad, 2001) and different choices of predictors (Benestad, 2004b; Linderson *et al*., 2004) indicate that the results may be sensitive to choices made in downscaling. Here, we will nevertheless use one fixed predictor domain in the comparison between different types of statistical precipitation models.

We will use the term 'model' here when referring to (1) GCMs, (2) the empirical models for downscaling analysis in general, (3) to discuss a particular type of downscaling model (precipitation model), (4) trend model, and (5) statistical regression models. We will distinguish between these by referring to (1) as 'GCMs', the downscaling models (2) as 'DSM', the precipitation models (3) as 'PM', and use the terms 'trend model' and 'regression models' for the two latter definitions. The

* Correspondence to: R. E. Benestad, The Norwegian Meteorological Institute, 0313 Oslo, Norway. E-mail: rasmus.benestad@met.no

notation *PM* will henceforth be used when referring to a particular type of DSM using either large-scale precipitation, sea-level pressure (SLP) or mixed-fields as predictor. We also make a distinction between 'station observation trend' estimates and 'downscaled trend' estimates.

Evaluation of downscaling techniques is problematic, as one DSM that works nicely in the present climate may not necessarily be capable of describing a changing climate. However, when sufficiently long precipitation series and gridded large-scale predictors are available, it is possible to calibrate the DSM on (the dependent) part of the data and use the rest of the (independent) data to test whether the DSM captures the observed historical precipitation variation and long-term changes. Such tests can provide some indication of the DSM's capability to describe long-term trends. It is furthermore possible to test the modelled relationship between large-scale features and local precipitation by taking one grid-box value from a GCM to represent the local climate (predictand), and use the large-scale anomalies representing several grid-box values from the same GCM over a given domain as predictor (Benestad, 2001). In the present study, DSMs trained with data from a 'calibration period' were evaluated by application over an independent 'validation period', and the validation not only involved correlation analyses, but also linear trend estimates.

Benestad (2004b) reported a systematic difference in downscaling based on large-scale SLP and precipitation, where the former did not produce any clear future trends while the latter suggested a shift in the mean precipitation as a consequence of global warming. Hence the motivation for using large-scale precipitation as a predictor for downscaling of local rainfall. Similar predictor-dependent differences between projected future precipitation trends are found in other studies, and reflect the fact that inclusion of at least one humidity related predictor is essential in order to capture the climate change associated with global warming, regardless of whether SLP is sufficient to account for inter-annual variability during a limited period (von Storch *et al*., 2000). One explanation is that SLP does not give a sufficient representation of changes in the atmospheric composition (i.e. moisture).

The main objective of this paper is to assess the choice of different predictor variables in the DSM, i.e. choice of PM, in terms of reproducing past precipitation trends as well as year-to-year fluctuations. We also examine the degree of non-stationarity in the statistical relationship between large and local scales. The paper describes the methodology in the 'Data & Methods' section and the main findings in the 'Results' section.

DATA AND METHODS

The predictand was monthly precipitation from the Nordklim data set (Tuomenvirta *et al*., 2001) [www.smhi.se/hfa_coord/nordklim (element code '601').] and the Norwegian stations were updated with recent observations from the Norwegian Meteorological Institute. The location of the 27 stations used in this study are marked on the map in Figure 1. The station data spans 1890–2004 for

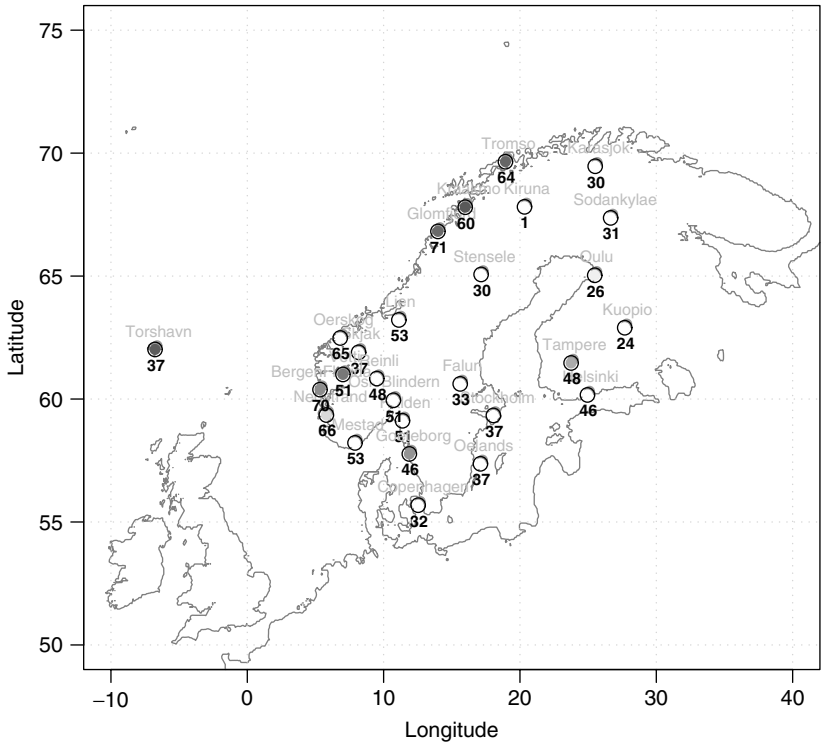


Figure 1. Map showing the locations of the stations. The location markers have different shadings, and the darker shading indicates high correlation (r) between the various trend estimates for the site. The numbers below the marks indicate the mean year-to-year correlation scores ('cor') based on all PMs and 'split-merge' as well as 'even' and 'odd' strategies.

DOI: 10.1002/joc

Norwegian stations whereas the other stations spanned the period 1890–1999.

The predictors were large-scale precipitation (see Benestad (2004b)) as well as SLP, and were taken from two data sources: (1) European Center for Medium Weather Forecast (ECMWF) re-analysis ERA40 (Simmons and Gibson, 2000) and from (2) National Center for Environmental Prediction (NCEP) re-analysis (Kalnay *et al*., 1996) obtained from NOAA CDC [http://www.cdc.noaa.gov/]. The reason for choosing these as predictors is that they also are GCM-generated, albeit through assimilation of observations, and hence bear similarities to precipitation from GCMs. Furthermore, there are not many other sources of reliable gridded precipitation available that cover both land areas as well as oceans. A third predictor, henceforth referred to as the 'mixed-field' predictor, was synthesised from a combination of the gridded large-scale precipitation and the SLP data (Benestad, 2004b). The synthesis of the mixed-field predictor is described in more detail in the Appendix. Both the precipitation and SLP fields were standardised by the standard deviation for the whole grid before being combined into one data set (i.e. the same scaling for all the grid-box values for a given variable). The predictor domain was 10 °W–40 °E/50 °N–75 °N and the time interval for the entire downscaling exercise was taken to be 1957–2002.

For most of the analysis presented here, the ERA40 and NCEP data have been used to downscale the local precipitation. However, in order to test the stationarity between large and small spatial scales, data from the ECHAM5 GCM (Roeckner *et al*., 1992; Oberhuber, 1993) from the Max-Planck-institute for Meteorology (following the IPCC SRES A1b scenario) was also used (obtained from the Program for Climate Model Diagnosis and Intercomparison (PCMDI) IPCC Internet site). The GCM simulation used here represents the future (2000–2200), and therefore all the following discussion concerning the past involves reanalyses and all the discussion concerning the 2000–2200 period refers to GCM data.

A similar common empirical orthogonal function (EOF) framework as described in Benestad (2001) was used for empirical-statistical downscaling, using the downscaling tool in the R-package `clim.pact` (Benestad, 2004a) [the version used here: clim.pact_2.2-0.tar.gz]. The downscaling involved a multiple regression analysis and was carried out using the '`DS`'-function (see the Appendix for details). The mixed-field predictor was constructed using the `clim.pact`-function '`mixFields`'. The downscaling was applied to monthly mean values of each of the 12 calendar months, and the seasonal values were subsequently obtained by taking the mean values of December, January, February (DJF) for winter, March, April, May (MAM) for spring, June, July, August (JJA) for summer, and September, October, November (SON) for autumn.

Three different strategies ('split-merge', 'even years', and 'odd years') were used for the downscaling of precipitation and studying the predicted trends in order to assess the robustness of the trend estimates with respect to arbitrary choices. The 'split-merge' strategy involved two stages: (1) first the data from 1957–1975 was used to calibrate the DSM, which was then used to make a prediction for the period 1976–2002; (2) a second DSM was then developed on the basis of the data from 1985 to 2002, and subsequently used to make predictions for the 1957–1984 period. The predictions for the two time periods were combined and compared with the observations. The two parts of the series were combined by enforcing the mean value for the overlap period (1976–1984) in the second sequence to be the same as in the first sequence in order to ensure that the two segments are combined with correct constant levels (described in more detail in the Appendix). Figure 2 shows the results of the 'split-merge' analysis, and since the downscaled results consist of independent data, a comparison with the results will not be subject to artificial skill. The merging of the two segments does introduce some uncertainty with respect to the subsequent trend analysis, as a preliminary analysis on series combined with no overlapping periods and no adjustments (not shown) gave significantly different trend estimates. Therefore, two additional sets of analyses were conducted. One set of analysis involved using even years to calibrate the DSMs and computing the trends over the predicted values for the odd years ('even years'), and the final set of analysis used odd year for calibration and even years for prediction ('odd years').

Benestad (2003) argued that there is no a priori reason a linear trend model gives a better description of past climate evolution than more complex trend models, and suggested the use of a polynomial fit as a more complex trend model to a time series. Past wintertime temperature evolution in northern Scandinavia indicates high temperatures in the 1930s followed by a period of cooling and then again a strong warming, which can be described in terms of a cubic equation. Higher order polynomials can also be used to account for decadal variations in precipitation over long intervals, as seen in Figure 2, but are unsuitable to describe shorter time series, such as 1957–2002, when strong fluctuations are present. The historical long-term evolution in the seasonal precipitation in Figure 2 does not follow a linear rate, but exhibits clear decadal variations. Over shorter intervals, a linear trend fit may nevertheless be justified as a means of providing an approximate change over the given period. In the present study, linear trend estimates will be evaluated as a measure for the change in precipitation, and it is important to note that the trend was estimated for exactly the same years in the downscaled analysis as for the station data when comparing these (typically over the 1957–1999 period). Identical intervals were also used when comparing series derived from ERA40 and NCEP.

Two different types of correlation analyses were applied: (1) correlation between downscaled time series and corresponding independent observations (a measure of how well the DSMs capture the year-to-year variations, i.e. as seen in Figure 2, listed in Table II,

DOI: 10.1002/joc

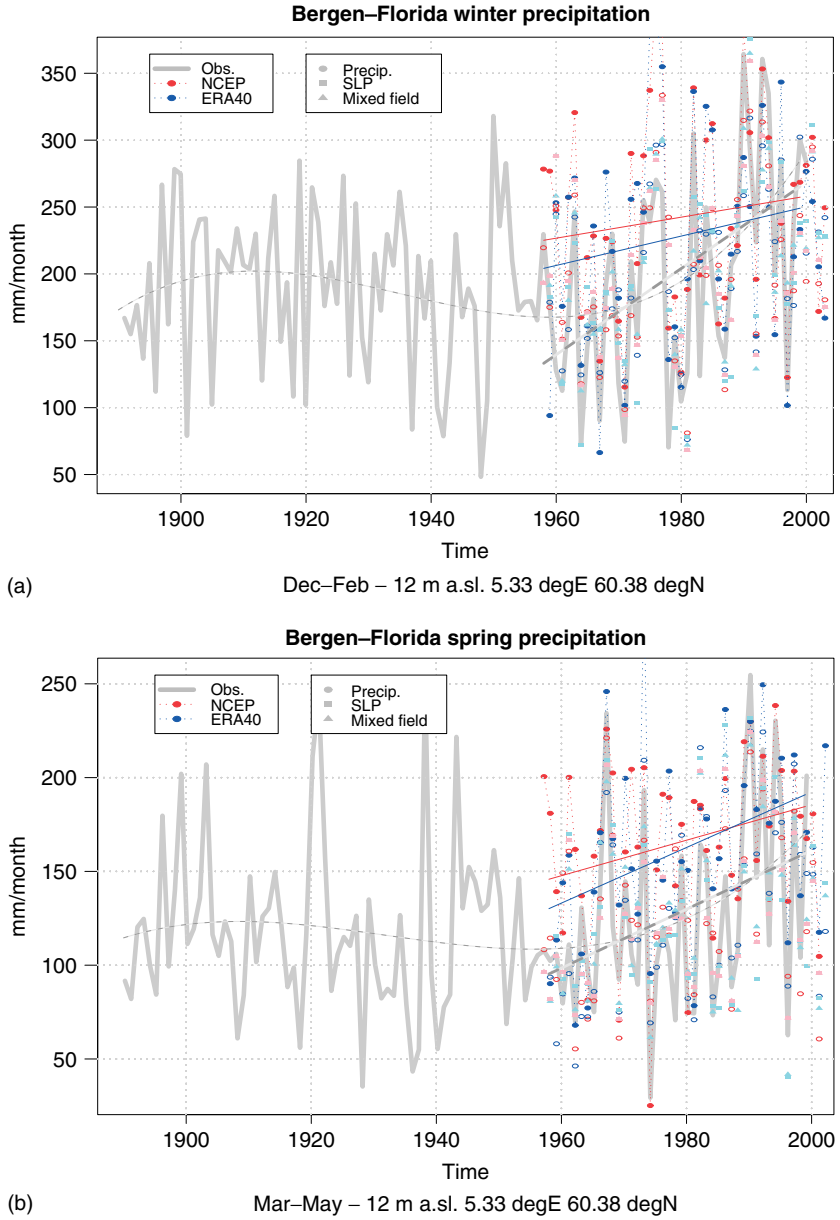


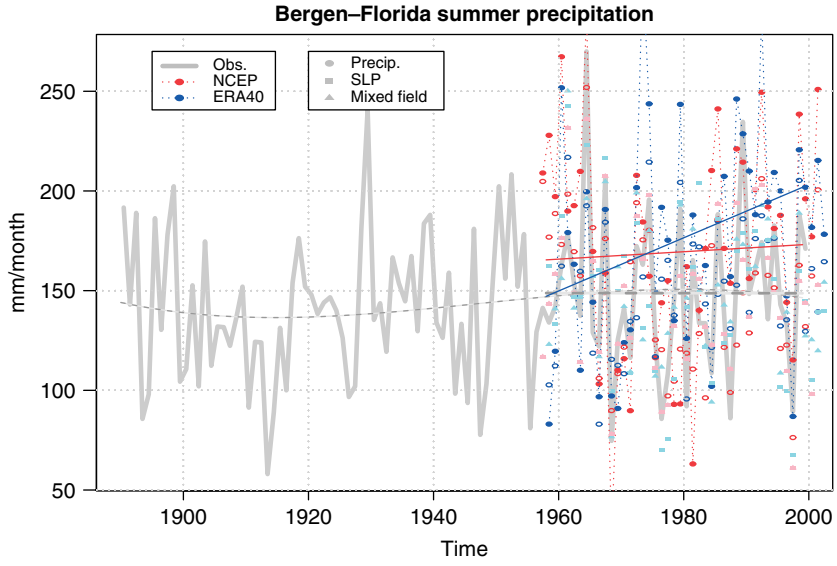
Figure 2. Time series showing the observed and 'split-merge' downscaled trends in precipitation in Bergen for the winter, spring, summer, and autumn seasons respectively. The blue and light blue symbols represent ERA40 results whereas the red and pink represent NCEP values. Circular symbols represent precipitation PMs, with filled symbols for 'split-merge' and open circles for odd and even. The square symbols mark results from SLP PMs, and triangles mark results from mixed-field PMs. Lines are only shown for 'split-merge' precipitation PMs, and only results from 'odd' and 'even' are shown for SLP and mixed-fields to avoid cluttering of the figures.

henceforth referred to as 'cor'); (2) correlation between observed and downscaled linear trend rates over the different stations (henceforth referred to as 'r') taking $r = \sum_i (x_i\, y_i)/\sqrt{\sum_i (x_i^2) \sum_i (y_i^2)}$ for observed (x) and downscaled trends (y) and without subtracting the mean values, since their absolute values are just as important in this case as the station-to-station variations.

RESULTS

A visual comparison between observed and downscaled precipitation in Figure 2 suggests that the DSMs manage to reproduce much of the year-to-year variance in the seasonal precipitation in Bergen on the west coast of Norway. The rainfall in Bergen is higher than for most of the other locations owing to the vicinity of the sea, prevailing westerlies, and orographic forcing. Table I shows the R^2-scores from the regression and hence the amount of the variance that the DSM can reproduce for different choices of PMs, different data sources, and seasons. The precipitation-based PM for Bergen rainfall achieves some of the highest scores. The R^2 statistic from a regression analysis indicates how much variance the DSM can account for, and the *adjusted* R^2 statistic presented here is a similar measure, but penalises for large number of predictors. Here, the adjusted R^2 is computed for each respective month by taking the

DOI: 10.1002/joc



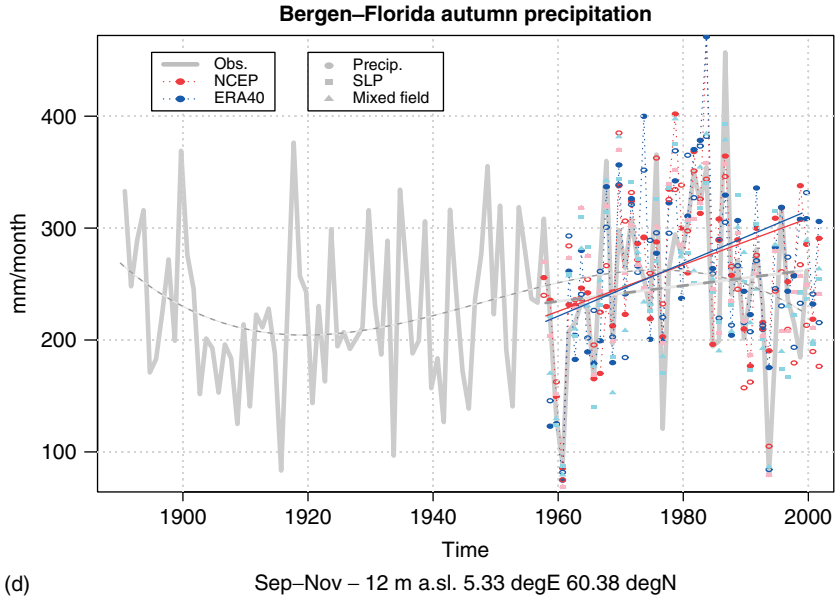


Figure 2. (*Continued*).

mean of the two values for each of the segments, and the range of values listed in Table I is subsequently estimated for the three months in each season. While the adjusted R^2 statistics in Table I give an indication of how close the best-fit is for the *dependent* data, Table II presents correlation scores (cor) giving an indication of how well the PMs reproduce the seasonal year-to-year variations over the 1957–1999 period for the prediction of *independent* data. Again, the PMs for Bergen score high for the 'cor'-score.

Table I indicates high skill scores in terms of R^2 statistics for all locations in general, but exceptions are found at stations in dry regions (e.g. Skjåk and Karasjok) and some stations close to the domain border (e.g. Torshavn). The scores are also, in general, lower in spring and summer than in autumn and winter. Table II and Figure 1 indicate that the skill scores in terms of inter-annual correlation over independent data are considerably less satisfactory for stations in the eastern part of the area, while the western stations tend to show skill scores closer to those based upon dependent data. This may suggest that the orographic effects of the more mountainous terrain in the western part of the area contribute to more well-defined links between the large-scale predictor fields and local precipitation. The lower scores inland (Figure 1) are probably due to more local-scale convective precipitation in the eastern and interior areas, while the precipitation in the western part is strongly influenced by large-scale orographic enhancement, and thus exhibits more well-defined links between large-scale prediction fields and local precipitation. The exceptional low 'cor'-score for Kiruna (1%) contrasts with the R^2 (>0.55) obtained in the regression analysis, as well as 'cor'-scores for the other stations. This low score may partly be due to

DOI: 10.1002/joc

Table I. Adjusted R^2-statistics from the regression analysis for a selection of stations and the 'split-merge' analysis. The range is based on estimates for the three months in each season.

Location	Predictor	DJF ERA40	MAM	JJA	SON	DJF NCEP	MAM	JJA	SON
Bergen	prec	75–87	78–91	63–80	72–88	81–93	66–88	65–76	69–91
	slp	78–81	70–88	77–85	75–90	74–81	72–91	73–86	76–90
	mix	79–87	69–90	79–88	77–91	81–89	74–90	74–85	70–94
Falun	prec	60–75	56–74	43–80	67–77	77–86	46–71	23–71	61–76
	slp	55–88	42–68	26–72	56–70	57–88	31–69	23–63	57–72
	mix	56–78	36–65	29–74	57–69	63–81	31–70	18–85	56–77
Glomfjord	prec	78–88	47–84	48–77	76–86	71–82	46–82	47–74	62–81
	slp	87–89	54–81	70–77	73–85	83–89	51–83	71–82	72–85
	mix	84–90	54–80	66–75	72–85	81–87	47–83	63–76	67–86
Göteborg	prec	67–67	68–86	64–76	61–76	72–83	44–86	38–66	68–80
	slp	64–67	40–87	48–63	52–68	57–75	46–85	46–58	51–66
	mix	63–75	55–85	46–64	66–72	63–68	30–86	40–63	63–72
Halden	prec	70–79	54–79	69–76	58–82	80–87	42–73	68–82	69–79
	slp	62–72	49–74	50–76	56–83	60–76	48–73	47–73	57–85
	mix	75–82	48–80	65–77	60–79	75–82	48–70	55–78	57–84
Helsinki	prec	71–87	51–75	49–80	56–65	70–85	56–71	39–68	73–85
	slp	68–74	38–75	55–55	47–75	57–80	38–64	48–64	44–71
	mix	75–84	46–66	58–79	52–72	59–84	49–58	54–68	69–82
Karasjok	prec	16–72	31–70	41–64	29–70	24–81	22–64	38–61	40–46
	slp	30–69	17–66	16–42	56–61	26–67	13–58	15–45	55–64
	mix	14–83	17–68	30–61	42–68	29–81	19–55	35–64	47–60
Kiruna	prec	78–92	82–83	66–86	61–91	66–92	66–74	69–83	55–85
	slp	74–78	72–91	62–71	80–88	73–79	70–85	54–70	79–86
	mix	69–84	72–82	71–84	71–87	76–90	55–77	63–89	81–86
Copenhagen	prec	58–84	75–80	61–78	59–75	53–85	57–67	38–75	54–84
	slp	44–83	59–69	52–81	56–83	31–87	58–72	48–73	61–86
	mix	50–76	51–72	56–79	62–72	40–75	57–72	37–63	64–81
Kråkmo	prec	66–83	47–78	43–83	65–85	51–83	48–77	46–84	67–84
	slp	73–84	51–67	55–79	70–79	69–82	43–68	56–78	69–81
	mix	66–88	54–83	68–74	71–82	62–80	51–80	52–73	68–89
Kuopio	prec	61–83	11–71	53–69	49–79	46–72	37–80	44–69	50–82
	slp	66–73	39–69	41–55	56–79	64–72	44–70	38–49	52–80
	mix	66–84	17–80	49–71	52–75	54–86	30–78	48–58	50–85
Lien	prec	61–88	40–73	31–60	70–75	62–73	36–88	32–51	74–84
	slp	58–84	64–70	43–60	71–82	58–81	62–72	44–63	67–84
	mix	56–91	47–75	30–61	70–81	62–77	54–84	26–59	69–86
Mestad	prec	76–86	66–81	64–82	74–89	74–82	65–75	51–82	78–87
	slp	70–76	66–80	46–65	74–84	71–75	70–84	50–67	75–85
	mix	78–81	67–78	51–76	65–85	77–82	64–77	56–63	76–86
Nedstrand	prec	64–81	69–89	44–66	73–92	74–83	65–84	40–79	75–95
	slp	69–75	75–81	67–80	72–87	71–73	73–82	70–81	71–87
	mix	70–80	81–84	62–81	71–90	75–83	77–82	62–86	72–93
Oslo	prec	58–87	46–78	37–70	50–83	74–83	48–67	50–73	59–85
	slp	54–85	48–62	42–69	59–66	53–85	54–63	49–73	65–70
	mix	59–89	45–67	46–73	52–75	63–81	46–66	49–77	48–83
Oulu	prec	58–81	54–79	70–78	54–89	54–68	58–76	63–71	56–88
	slp	52–70	64–83	41–67	52–83	55–77	60–88	47–69	55–83
	mix	34–75	63–87	34–64	54–90	63–88	60–65	53–67	56–91
Reinli	prec	59–74	42–74	48–66	57–76	56–78	47–68	60–81	57–72
	slp	60–72	45–59	22–50	46–66	58–78	52–60	24–54	48–65
	mix	59–73	41–61	33–53	57–74	61–83	41–58	42–65	58–70
Skjåk	prec	56–74	44–75	48–60	18–59	52–80	41–74	47–59	60–63
	slp	49–69	22–69	32–64	44–63	46–71	17–65	27–65	45–65
	mix	52–71	40–70	42–67	32–67	48–62	42–72	40–53	34–64
Sodankylä	prec	44–63	18–65	38–75	48–78	52–65	30–59	43–73	52–74
	slp	56–81	56–66	31–59	57–73	55–78	51–58	29–62	58–75
	mix	41–78	57–70	24–74	55–71	52–84	35–62	53–63	63–83

DOI: 10.1002/joc

Table I. (*Continued*).

Location	Predictor	DJF ERA40	MAM	JJA	SON	DJF NCEP	MAM	JJA	SON
Stensele	prec	58–69	24–59	56–75	18–81	47–72	35–60	60–70	23–76
	slp	63–75	46–61	64–78	46–89	65–73	38–61	65–72	46–87
	mix	63–72	51–65	61–73	39–85	66–74	46–62	57–75	40–84
Stockholm	prec	72–88	49–57	43–82	52–90	72–91	46–55	36–66	55–60
	slp	53–86	38–55	28–56	56–83	42–87	41–56	26–50	57–86
	mix	69–90	38–46	19–72	60–94	66–87	29–57	43–58	56–84
Tampere	prec	65–81	23–68	63–88	38–61	62–69	22–62	43–77	41–70
	slp	63–69	27–66	41–75	48–66	62–72	26–64	45–74	47–65
	mix	65–71	23–63	68–80	51–71	59–74	17–67	65–79	64–72
Torshavn	prec	53–73	61–78	−3–64	52–78	59–77	30–71	0–57	31–83
	slp	58–74	44–65	39–41	51–78	45–72	44–55	33–42	49–77
	mix	60–76	43–70	19–57	49–74	56–71	39–50	22–62	53–73
Tromsø	prec	73–82	62–80	58–62	60–82	69–79	54–81	47–68	59–86
	slp	73–89	55–79	49–81	48–85	73–88	62–75	55–78	47–86
	mix	75–88	75–78	54–71	66–87	77–85	75–78	48–71	58–87

Table II. Correlation coefficients estimated for the independent data in the 'split-merge' analysis. The correlation scores shown here are the mean of the two independent periods.

Location	Predictor	DJF ERA40	MAM	JJA	SON	DJF NCEP	MAM	JJA	SON
Bergen	prec	0.71	0.77	0.59	0.7	0.83	0.71	0.52	0.67
	slp	0.76	0.67	0.73	0.66	0.68	0.59	0.66	0.66
	mix	0.73	0.81	0.74	0.66	0.78	0.77	0.7	0.71
Falun	prec	0.58	0.15	0.23	0.12	0.48	0.43	0.47	0.15
	slp	0.49	0.34	0.37	0.2	0.41	0.39	0.32	0.18
	mix	0.48	0.18	0.12	0.31	0.56	0.39	0.35	0.27
Glomfjord	prec	0.83	0.69	0.4	0.65	0.79	0.65	0.43	0.67
	slp	0.77	0.66	0.72	0.83	0.79	0.59	0.68	0.83
	mix	0.85	0.77	0.54	0.79	0.8	0.8	0.58	0.84
Göteborg	prec	0.45	0.2	0.57	0.44	0.53	0.49	0.5	0.53
	slp	0.44	0.61	0.44	0.37	0.45	0.52	0.35	0.37
	mix	0.51	0.58	0.39	0.42	0.42	0.53	0.47	0.55
Halden	prec	0.74	0.6	0.47	0.58	0.58	0.61	0.28	0.61
	slp	0.45	0.55	0.52	0.29	0.5	0.51	0.45	0.29
	mix	0.52	0.48	0.52	0.49	0.47	0.61	0.5	0.58
Helsinki	prec	0.66	0.4	0.35	0.69	0.69	0.15	0.22	0.47
	slp	0.5	0.38	0.41	0.44	0.53	0.4	0.48	0.45
	mix	0.54	0.3	0.39	0.66	0.67	0.29	0.33	0.56
Karasjok	prec	0.5	0.44	0.29	0.2	0.13	0.49	0.37	0.13
	slp	0.43	0.44	0.37	0.18	0.47	0.29	0.1	0.2
	mix	0.46	0.14	0.22	0.32	0.2	0.2	0.32	0.31
Kiruna	prec	0.4	0.09	−0.05	0.01	0.07	−0.02	−0.32	0.02
	slp	−0.01	−0.14	0.5	−0.42	−0.02	−0.14	0.55	−0.39
	mix	0.05	−0.24	0.32	−0.2	0.01	−0.07	0.1	0.09
Copenhagen	prec	0.41	0.46	0.21	−0.01	0.48	0.41	0.45	0.14
	slp	0.16	0.32	0.55	0.26	0.11	0.31	0.51	0.32
	mix	0.3	0.19	0.36	0.38	0.36	0.6	0.27	0.17
Kråkmo	prec	0.83	0.51	0.27	0.41	0.69	0.47	0.23	0.71
	slp	0.74	0.59	0.71	0.78	0.7	0.53	0.62	0.75
	mix	0.79	0.5	0.36	0.69	0.74	0.52	0.49	0.66
Kuopio	prec	0.5	0.09	−0.14	0.48	0.38	−0.04	0.23	0.42
	slp	0.51	0.12	0.18	0.03	0.55	0.01	0.16	0.17
	mix	0.57	0.24	−0.07	0.21	0.5	0.05	0.1	0.45

(*continued overleaf*)

DOI: 10.1002/joc

Table II. (*Continued*).

Location	Predictor	DJF ERA40	MAM	JJA	SON	DJF NCEP	MAM	JJA	SON
Lien	prec	0.71	0.42	−0.08	0.68	0.7	0.61	−0.06	0.64
	slp	0.75	0.61	0.14	0.7	0.73	0.6	0.12	0.72
	mix	0.8	0.66	0.12	0.67	0.76	0.64	0.2	0.78
Mestad	prec	0.67	0.63	0.43	0.68	0.77	0.63	0.47	0.63
	slp	0.42	0.49	0.5	0.37	0.56	0.49	0.36	0.39
	mix	0.64	0.61	0.48	0.61	0.47	0.65	0.3	0.56
Nedstrand	prec	0.71	0.57	0.42	0.76	0.78	0.72	0.5	0.68
	slp	0.72	0.75	0.44	0.61	0.73	0.69	0.62	0.64
	mix	0.66	0.71	0.65	0.74	0.78	0.73	0.52	0.73
Oslo	prec	0.63	0.71	0.54	0.41	0.57	0.26	0.38	0.71
	slp	0.27	0.64	0.55	0.42	0.23	0.56	0.61	0.43
	mix	0.52	0.55	0.59	0.53	0.55	0.49	0.33	0.68
Oulu	prec	0.33	0.54	0.45	0.02	0.4	0.32	0.45	0.02
	slp	0.53	0.17	0.28	0.22	0.3	0.09	0.27	0.22
	mix	0.52	0.01	0.41	0.01	0.51	0.07	0.26	−0.04
Reinli	prec	0.56	0.39	0.36	0.31	0.6	0.41	0.4	0.58
	slp	0.55	0.64	0.56	0.45	0.51	0.58	0.47	0.4
	mix	0.64	0.56	0.34	0.37	0.47	0.43	0.49	0.55
Skjåk	prec	0.46	0.24	0.03	0.43	0.65	0.39	0.1	0.08
	slp	0.65	0.53	0.16	0.28	0.49	0.57	0.3	0.28
	mix	0.66	0.4	0.08	0.3	0.78	0.45	0.21	0.26
Sodankylä	prec	0.47	0.21	0.25	0.31	0.63	−0.02	0.35	0.37
	slp	0.63	0.15	0.17	0.21	0.55	0.14	0.23	0.15
	mix	0.6	0.24	0.41	0.21	0.57	0.02	0.27	0.37
Stensele	prec	0.63	0.37	0.27	0.23	0.51	0.37	0.32	−0.03
	slp	0.46	0.49	0.39	−0.27	0.59	0.46	0.43	−0.27
	mix	0.69	0.34	0.35	−0.07	0.35	0.37	0.34	−0.15
Stockholm	prec	0.62	0.09	0	0.17	0.61	0.31	0.34	0.28
	slp	0.59	0.32	0.27	0.51	0.43	0.24	0.52	0.48
	mix	0.48	0.11	0.13	0.56	0.61	0.39	0.41	0.4
Tampere	prec	0.64	0.47	0.55	0.61	0.57	0.31	0.37	0.63
	slp	0.44	0.15	0.52	0.51	0.48	0.1	0.6	0.51
	mix	0.64	0.22	0.69	0.52	0.51	0.47	0.53	0.58
Torshavn	prec	0.6	0.29	0.36	0.27	0.54	0.16	0.01	0.2
	slp	0.31	0.39	0.21	0.46	0.38	0.49	0.33	0.46
	mix	0.52	0.48	0.46	0.3	0.58	0.54	0.14	0.29
Tromsø	prec	0.74	0.66	0.43	0.58	0.77	0.53	0.34	0.63
	slp	0.85	0.68	0.57	0.78	0.84	0.56	0.58	0.74
	mix	0.82	0.69	0.51	0.65	0.85	0.56	0.42	0.65

the fact that the station data for Kiruna stopped in 1990 and therefore only 15 data points in the second split-merge sequence were available for the calibration of the DSM for 1976–1999.

An important question is whether there are significant differences between the downscaled results derived from the two re-analysis products or between the different PMs. A comparison between adjusted R^2 statistics associated with the two data sources (not shown) indicates that the two reanalyses give similar DSM skill (peaking at $R^2 \sim 60$–80%). The evaluation of the correlation scores (not shown) suggests somewhat greater differences than for the adjusted R^2 statistic, but there is still an indication that ERA40 yields similar skill as the NCEP re-analysis. Similar comparisons of R^2-estimates from different PM types show that the choice of predictor may be slightly more important than the data source (not shown), and an evaluation of the three different PM types in terms of the 'cor'-score (not shown) shows similar trends in general, with correlations peaking at cor $\sim$0.3–0.7 and a hint that the mixed-field and the SLP PMs are slightly superior to the precipitation-based PM for the most skillful locations. In summary, all the PMs suggest a reasonable skill in reproducing the local year-to-year variations.

Figure 3 presents a graphical overview of observed and downscaled linear trend slope estimates for Bergen. The graphics shows 18 different values for each season (winter–autumn from left to right), and these 18 values are divided into two sets representing ERA40 (blue shaded symbols) and NCEP (red shaded symbols), for each of which there are nine estimates. Each subset of nine is then divided into three groups-of-three subsequent

DOI: 10.1002/joc

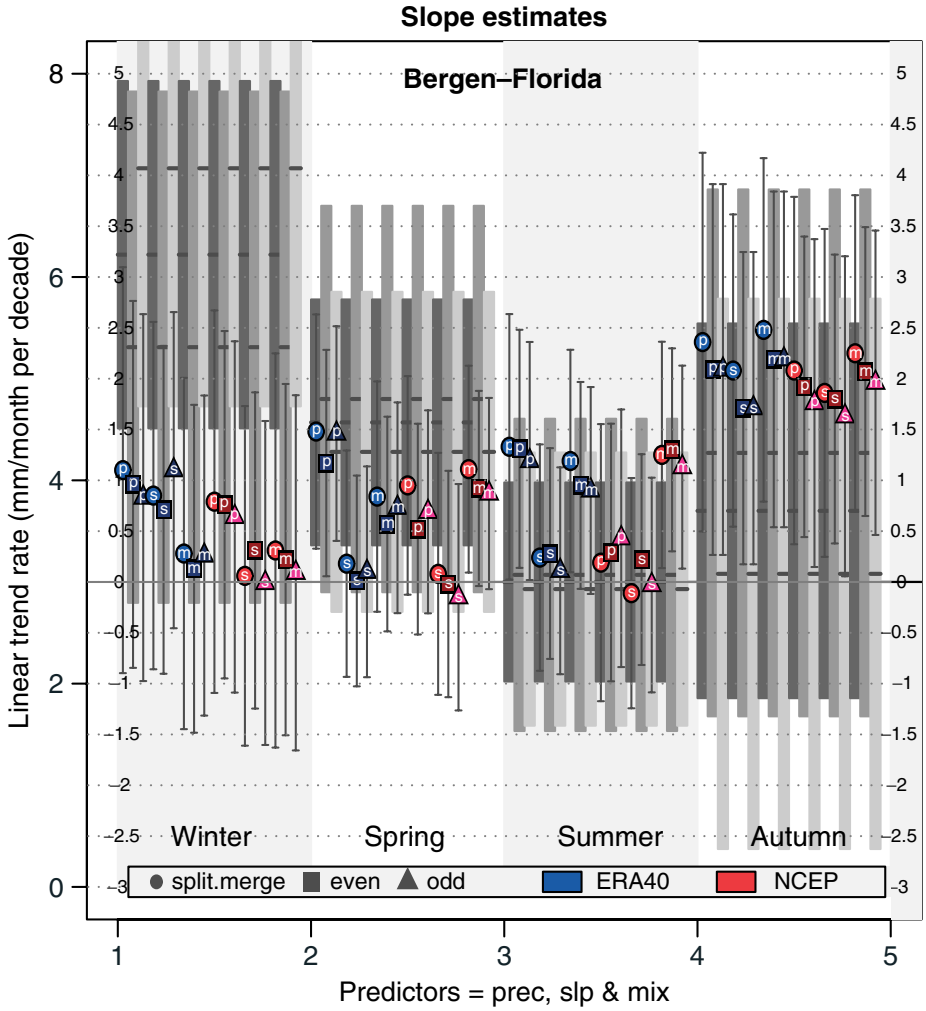


Figure 3. A graphical presentation for Bergen showing the estimated linear trend slope using different predictors (precipitation, SLP, and mixed represented by groups of 3 points shown in the same order and marked with 'p', 's', and 'm', respectively) using 'split-merge', 'odd years', and 'even years' (marked with different symbols), and derived using different data sets (ERA40 in blue and NCEP in red). The different seasons are presented as winter, spring, summer, and autumn from left to right. The error bars indicate the 95% confidence interval of the trend estimate. Thick grey vertical background columns indicate the trend estimates for the station series and the shaded areas show corresponding 95% confidence interval. The downscaled and observed trends were calculated for the same years.

data points: each point in a group (marked with circle, square, and a triangle) shows the results for the same PM but for the 'split-merge', 'odd years', and 'even years' strategies, respectively. The three 3-point groups represent the results for three PM types and are marked with the symbols 'p', 's', and 'm' for precipitation, SLP, and mixed-field, respectively. Thus, Figure 3 provides a comparison between the results from 'split-merge', 'odd years', and 'even years' strategies as well as between different PMs and calibration data sets.

In general, there is a smaller scatter between 'split-merge', 'odd years', and 'even years' strategies for one given predictor (intra-predictor scatter) than between different predictors. The mixed-field PMs (symbols marked with 'm') tend to yield similar trend estimates for the two reanalyses, but the ERA40 and NCEP SLP PMs also yield similar trends (symbols marked with 's'), with the exception of the winter season. The greatest differences between the two reanalyses can be seen in the results from the precipitation PMs (symbols marked with 'p'). Note that the summer station observation trend estimates for Bergen were not well defined, and that the downscaled trends were very sensitive to the PM used, as well as the interval used for the trend analysis (not shown). Furthermore, the results presented in Figure 3 suggest that the trend estimates for the station observations were not always robust with respect to the analysis strategy (i.e. regarding the years that were included in the analysis), as the 'split-merge', 'odd years', and 'even years' gave different trends for Bergen for the analysis based on the station observations (shaded columns in the background). The downscaled results, on the other hand (symbols with error bars), were not as sensitive to the choice of strategy. The choice of PM type, however, was more important, as for instance, a SLP-based PM calibrated with the ERA40 data gave weaker summer trends than the corresponding precipitation- or mixed-based PMs. For the PMs calibrated on the NCEP re-analysis, on the other

DOI: 10.1002/joc

hand, both the precipitation- and SLP-based PMs gave little trend in summer whereas the mixed-field PM indicated a positive trend.

Although the 95% confidence intervals of the observed and downscaled trends tend to overlap, it is evident that the trends estimated from the station observations were not well-captured by the downscaling. Figure 4 shows a comparison between observed and downscaled trends derived from the two data sources, where each point indicates the values for a given location. The correlations shown in the figure legend gives a measure of how well the DSMs reproduced the observed trends. Both the scatter of points in the plot and the correlation values support the impression from Figure 3 that the DSMs did not skillfully reproduce the linear precipitation trends over the 1957–2002 period. It is interesting to note the low correlation (∼0.5) between trend estimates from ERA40 and NCEP in Figure 4, which suggests that the trends, in general, are associated with a degree of uncertainty, as expected for short series with strong year-to-year variations. The two reanalyses used here are produced with different models that have different spatial resolution.

A discrepancy between trends derived from station series and re-analysis over the 1957–1999 period can also be inferred from Table III, which lists the trend estimates for interpolated, as opposed to *downscaled*, precipitation, as well as correlation and root-mean-square-error with respect to the station observations. Table III further shows a summary for the entire year, including the similarity between the annual cycle in the observations and the reanalyses. In general, the two reanalyses obtain similar 'cor'-scores, with the IQR estimates of 0.69–0.78 for ERA40 all-year month-to-month correlation and a corresponding 0.69–0.79 for NCEP. However, their interpolated series yield quite different trends, especially in NCEP (here in %/century). On an annual basis, 6 stations (Skjåk, Bergen, Vetti, Ørskog, Torshavn, and Göteborg) indicate a positive trend, which is significant at the 5% level in the station observations. All these localities show positive trends also in ERA40, and four of them are significant. In NCEP, on the other hand, four of these localities show negative trends, and none are significant. Further, NCEP indicated significant negative trends for 11 of the remaining stations that did not exhibit significant trends in the observations. A discrepancy between (the annual cycle of) the precipitation interpolated from NCEP re-analysis and GCMs and station series has also been noted in previous analysis (Benestad, 2004b).

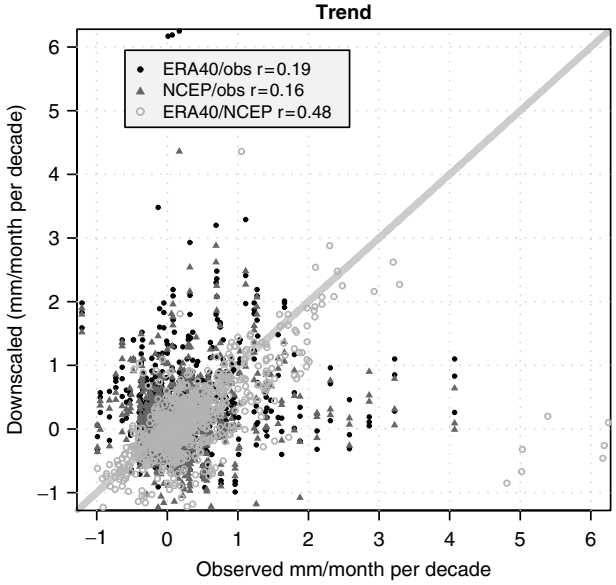


Figure 4. Comparison between observed and downscaled linear trend for different locations derived from the ERA40 and NCEP reanalyses. The correlation analysis shown in the legend was done for the observed and downscaled trend for the different stations, and is different to the correlation analysis shown in Table I. The comparison between ERA40 and NCEP (open circles) is shown as ERA40 values along the x-axis and NCEP along the y-axis. The analysis presents results from 'split-merge', 'odd years', and 'even years' strategies. The downscaled and observed trends were calculated for the same years.

Figure 5 shows the correlation between the DSM-derived and observed trends over the station network. The scatter plot suggests that the SLP-based PMs (shown as the middle set of three values in the groups-of-three in Figure 3) give the best reproduction ($r = 0.25$ as opposed to 0.19 and 0.13) of the linear precipitation trends. In Figure 6, the above analyses are stratified in terms of seasons as well as PM type in order to examine whether the DSM skill is seasonally dependent. There is a clear tendency towards higher correlation scores in winter and lower scores in summer (left panels) when convective rain-processes are more prevalent. None of the PM types give a good reproduction of the trends in summer and autumn (right panels). The correlation between the trends derived from SLP is 0.45 for the winter season whereas the correlation derived using the precipitation-based PM is 0.44 and the mixed-field 0.18, indicating strong sensitivity to the choice of predictor for calibrating the DSM in this season. The precipitation- and mixed-field PMs obtain similar scores as the SLP-based PM in spring ($r = 0.31$, $r = 0.37$, and $r = 0.29$, respectively). Some information on the geographical spread of skill can be obtained from Figure 1, in which the grey-scale of the location markers reflect the mean r skill score (darker shading implies higher r score indicating similar trend estimates) and the 'cor'-score for year-to-year agreement is given below the marks. The mean 'cor'-score is generally lower in northern Sweden and northern Finland, whereas the best reproduction of the past trends (highest r values) were found along the western and northwestern coast of Norway as well as on Torshavn, Göteborg, and Tampere. In summary, few of the downscaling exercises reproduced the linear trend in precipitation over the 1957–2002 period satisfactorily, despite relatively high R^2 and 'cor' correlation scores.

A well-known caveat associated with empirical downscaling is the risk of non-stationary relationships between the large-scale features used as predictors and the local response; it is important to test for this. Such

DOI: 10.1002/joc

Table III. Comparison between interpolated re-analysis data and station series for all calendar months of the year taken together in the period 1957–1999. The interpolated re-analysis precipitation has been scaled to have the same standard deviation (s) as the observations. The linear trend estimates in columns 1–3 are given as %/(100 year) $\pm 2 \times s$, and the *observed* mean over 1957–1999 was used as reference climatology in estimating the percentage for the reanalyses as well as observations. The values in the parentheses are the p-value in % (<5 for statistically significant at the 5% level). The two last columns list the correlation (cor) and the root-mean-square-error scores for ERA40 and NCEP respectively. The correlation r between percentage (%/year) trend estimates $r = 0.48$ between observation and ERA40, 0.44 between obs and NCEP, and 0.07 between ERA40 and NCEP.

Location	Obs. trend	ERA40 trend	NCEP trend	ERA40 cor/rmse	NCEP cor/rmse
Halden	14 ± 48 (54)	25 ± 44 (28)	-46 ± 44 (4)	0.76/1.28	0.82/1.11
Skjåk	60 ± 56 (3)	56 ± 56 (5)	-28 ± 56 (32)	0.61/0.78	0.63/0.76
Oslo-Blindern	-5 ± 44 (85)	27 ± 44 (25)	-45 ± 44 (5)	0.77/1.2	0.77/1.19
Reinli	-5 ± 50 (82)	39 ± 52 (14)	-47 ± 52 (7)	0.64/1.53	0.59/1.64
Mestad	1 ± 48 (95)	-1 ± 48 (98)	-13 ± 46 (57)	0.84/2.27	0.84/2.3
Nedstrand	37 ± 44 (10)	35 ± 44 (11)	-9 ± 42 (67)	0.77/2.76	0.81/2.53
Bergen-Florida	76 ± 44 (0)	48 ± 44 (3)	-1 ± 42 (96)	0.83/2.8	0.85/2.62
Vetti	86 ± 64 (1)	60 ± 62 (6)	0 ± 62 (99)	0.78/1.69	0.81/1.55
Ørskog	57 ± 46 (1)	47 ± 46 (4)	-22 ± 46 (34)	0.81/2.28	0.83/2.16
Lien	29 ± 40 (17)	-9 ± 44 (66)	-40 ± 40 (5)	0.54/1.72	0.72/1.32
Glomfjord	20 ± 50 (44)	-7 ± 50 (79)	-45 ± 48 (6)	0.83/3	0.84/2.92
Tromsø	43 ± 48 (7)	23 ± 46 (32)	-80 ± 46 (0)	0.81/1.38	0.62/1.93
Karasjok	3 ± 62 (91)	45 ± 56 (11)	-122 ± 56 (0)	0.66/0.88	0.77/0.72
Copenhagen	-16 ± 48 (52)	-28 ± 44 (21)	-26 ± 44 (21)	0.71/1.07	0.75/0.99
Helsinki	-5 ± 40 (81)	-12 ± 40 (52)	-16 ± 40 (44)	0.68/1.13	0.69/1.12
Tampere	46 ± 44 (4)	17 ± 44 (43)	-24 ± 44 (25)	0.71/0.95	0.75/0.89
Kuopio	31 ± 46 (16)	-31 ± 42 (14)	-48 ± 42 (2)	0.67/1.1	0.64/1.14
Torshavn	49 ± 38 (1)	51 ± 36 (1)	24 ± 36 (18)	0.73/1.97	0.73/1.98
Göteborg	74 ± 42 (0)	15 ± 42 (52)	-34 ± 42 (11)	0.74/1.2	0.77/1.15
Øland	0 ± 46 (100)	-37 ± 46 (11)	-3 ± 46 (89)	0.64/0.85	0.63/0.87
Sodankylä	-9 ± 42 (66)	-5 ± 46 (84)	-95 ± 42 (0)	0.75/0.8	0.67/0.89
Oulu	-13 ± 44 (53)	-20 ± 48 (43)	-138 ± 44 (0)	0.72/0.89	0.65/0.98
Stockholm	-4 ± 48 (88)	9 ± 48 (69)	-15 ± 44 (50)	0.74/0.95	0.74/0.95
Falun	-2 ± 46 (94)	10 ± 46 (65)	-46 ± 46 (4)	0.73/1.03	0.76/0.97
Stensele	29 ± 52 (29)	2 ± 52 (92)	-62 ± 52 (1)	0.77/0.93	0.74/0.98
Kiruna	-10 ± 74 (81)	51 ± 82 (21)	-141 ± 82 (0)	0.7/1.2	0.68/1.22
Kråkmo	43 ± 54 (11)	-5 ± 52 (82)	-59 ± 50 (2)	0.84/2.04	0.73/2.65

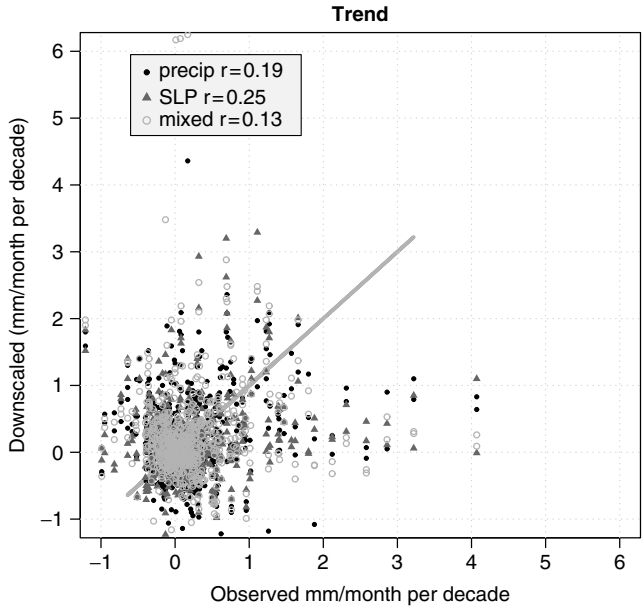


Figure 5. Comparison between observed and downscaled linear trend for the different locations derived using the predictors large-scale precipitation, SLP, and mixed-fields. The analysis presents results from 'split-merge', 'odd years', and 'even years' strategies. The downscaled and observed trends were calculated for the same years.

DOI: 10.1002/joc

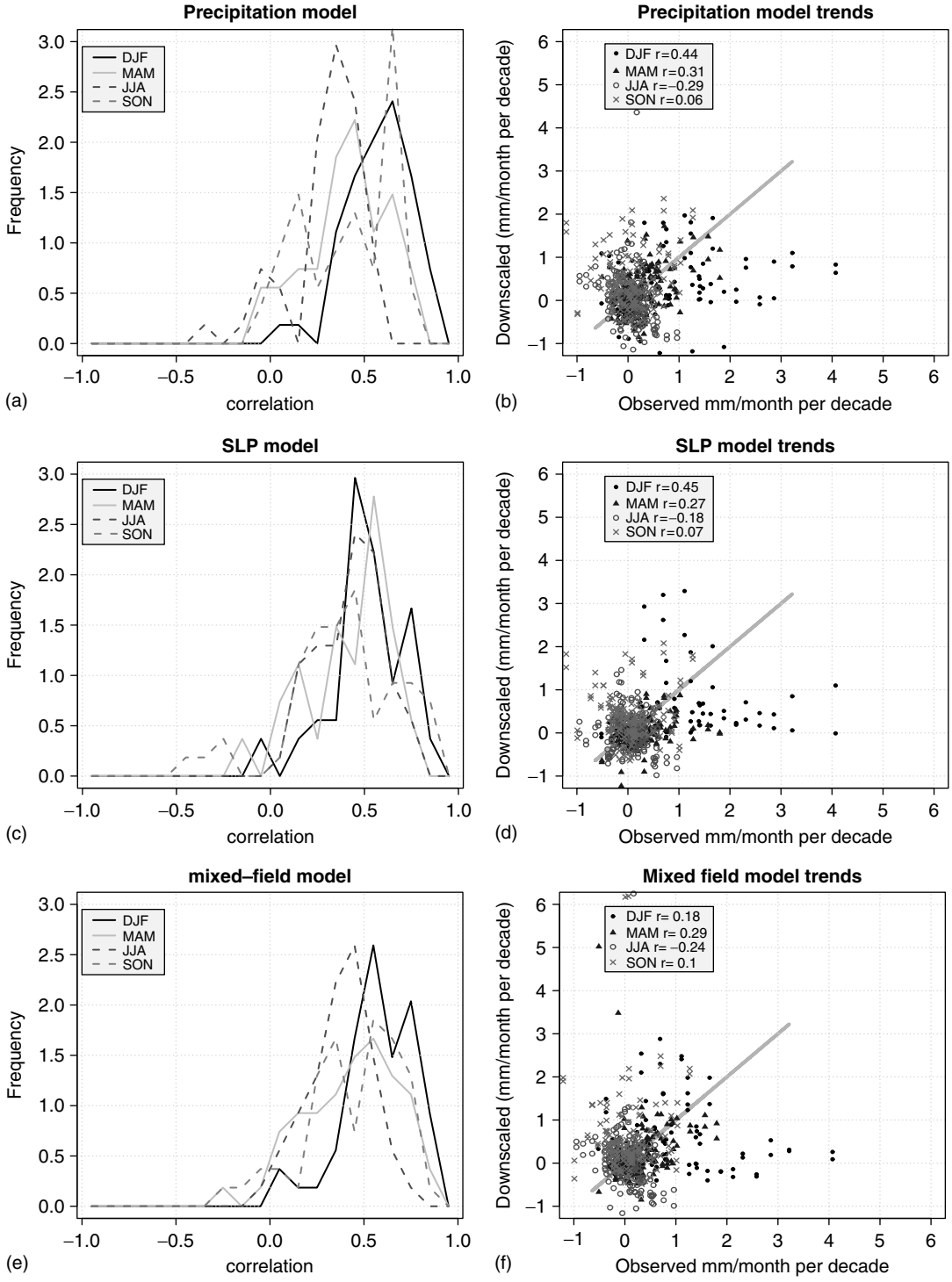


Figure 6. Evaluation of the different DSM skill and their dependency on season for PMs, using large-scale precipitation (a–b), SLP (c–d), and mixed-field (e–f) as predictors. The left panels show histograms of the correlation scores and the right panels show the comparison between observed and downscaled trends.

non-stationarities can involve a change in the relationship between the large-scale spatial structures and the local variations. The gridded precipitation data only go back to 1957, but gridded SLP products extend back to the nineteenth century. One severe limitation of the re-analysis data is their short extent back in time, and we therefore applied the analysis to SLP, which provides a longer gridded data record than the precipitation. Figure 7 shows a stationarity test based on SLP carried out for Bergen, based on the Det Norske Meteorologisk Institutt (DNMI) SLP (Benestad, 2000; Benestad and Melsom, 2002) (included in the `clim.pact` package [`'data(DNMI.slp)'`]). The data in the grey hatched region were used for the DSM calibration, whereas the remaining independent data were used for evaluation. The predictions (grey solid lines) for the independent periods

DOI: 10.1002/joc

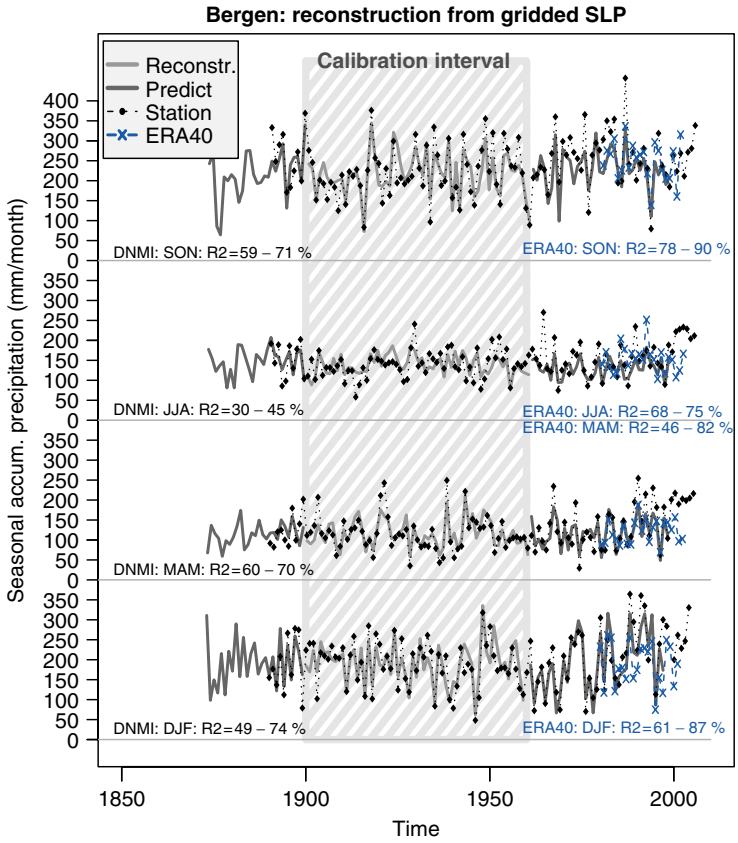


Figure 7. Reproduction (dependent analysis, model calibration, shown in the hatched interval) and independent prediction (outside the hatched area) of seasonal precipitation in Bergen using the DNMI SLP. Similar analysis based on ERA40 is also shown (dash & cross), although for the latter data set, the calibration period was 1957–1979 (i.e. shorter than the DNMI calibration). The panels show the four different seasons: winter (bottom), spring, summer, and autumn (top).

follow the 'truth' (dots) closely for most of the time, except for the most recent two decades in which the observed winter and spring precipitation appear to be systematically higher than the downscaled results. Also, the observed five most recent summer precipitation amounts are higher than the predictions. Otherwise, close tracking over time suggests that non-stationary relationships are not strongly present. The analysis was repeated with the ERA40 SLP over a shorter interval (blue, showing only *dependent* values here), which also exhibits a divergence from the most recent observed winter and spring precipitation. Hence, the high precipitation amounts during the seven most recent years do not appear to be related to just a change in the circulation; however, there is otherwise no clear sign of non-stationarity.

Figure 8 shows results from a downscaling analysis using large-scale precipitation from ECHAM5 (90 °W–60 °E/30 °N–80 °N) as predictor and taking interpolated precipitation from ECHAM5 (for same location as Bergen) as the predictand rather than station observations as in Figure 7. The calibration period was '2000'–'2100' (hatched region), and the '2100'–'2200' interval was used for independent validation. Thus, this exercise mimics the ordinary empirical downscaling exercise, but now using GCM data as predictor and predictand, both for calibration as well as for evaluation. This allows us to test the predictions against a 'true value', where the evaluation consisted of using large-scale precipitation anomalies from the GCM to predict the corresponding values for the independent second half of the data, and then subsequently compare them with the interpolations. Also, shown in the figure as crosses are corresponding results obtained with SLP-based PMs. The precipitation- and the SLP-based analysis produced similar results, albeit for the slight tendency for local precipitation being underestimated by the SLP during winter (autumn, top). The close agreement between the 'truth' and the precipitation- and SLP-based PMs suggests that there is no predicted change in the relationship between small and large spatial scales. Hence, the same large-scale conditions responsible for local climate variations during the calibration interval can be used to predict local variations at a later stage. Furthermore, the ECHAM5 simulation does not indicate any dramatic changes in the precipitation, and a substantial portion (The range of R^2 for the different calendar months was 45–69%) of the precipitation interpolated from the ECHAM5 GCM to Bergen can be predicted from either the SLP-field or the large-scale precipitation.

DISCUSSION AND CONCLUSION

Most of the past seasonal precipitation trends from station observations in Fennoscandia have been marginally

DOI: 10.1002/joc

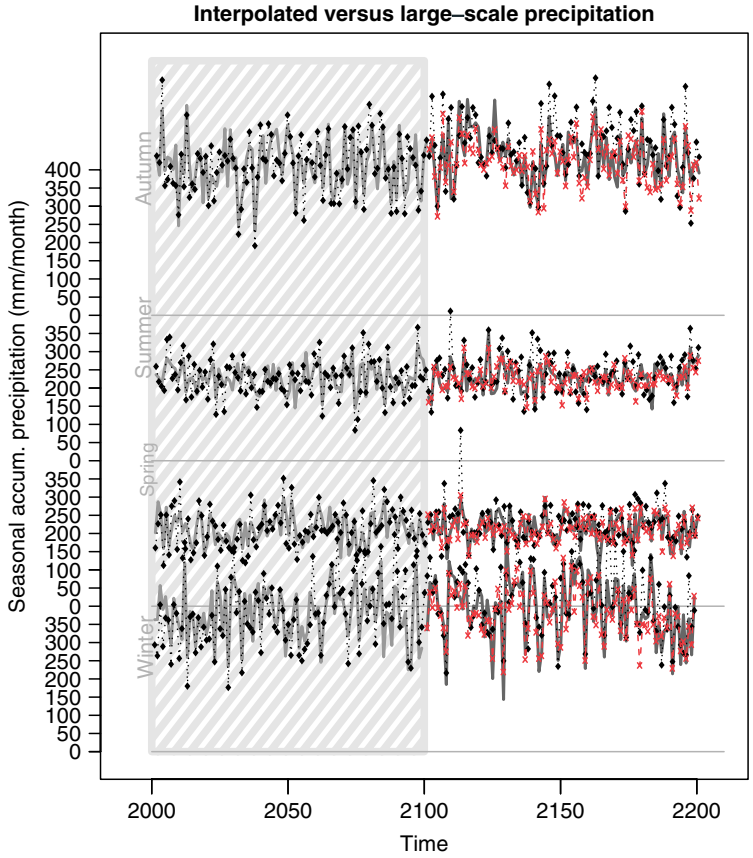


Figure 8. Test of non-stationarity of the relationship between local precipitation in Bergen (here interpolated from nearby grid values) and large-scale patterns (here EOFs): a 'reconstruction' from the downscaling analysis (grey line) is compared with the actual interpolated values (black symbols). The Panels show the four different seasons: winter (bottom), spring, summer, and autumn (top). The red lines/symbols show similar downscaling based on SLP instead of large-scale precipitation. The hatched area shows the calibration period and dependent data.

positive during the period 1957–2002, and only a few locations exhibited trends that were statistically significant at the 5% level. The trend analysis for the downscaled results did not reproduce the observed a 50-year trends well, partly because the trends often were not statistically significant or well defined.

For trend analysis on the entire set of stations and seasons (not only those with significant trends), the SLP-based PMs gave better agreement with the observed trends than the mixed-field and precipitation-based models. Seasonal analyses showed that the precipitation-based PMs were of equal quality in all seasons except in summer, while the mixed-field models were inferior also in winter. This result does not support the notion that models including a humidity related predictor will, in principle, be more able to capture trends than SLP-based PMs because they can represent changes in both circulation and air properties.

There are, however, three factors that can cause substantial degrees of uncertainty in the present trend analysis: (1) the downscaled and interpolated series were relatively short, with large year-to-year variations, reducing the signal-to-noise ratio in the trend analysis, and the trends were not well defined; (2) the precipitation in the reanalyses is derived from models rather than observations, and since air humidity is a poorly constrained quantity, the precipitation fields from the reanalyses often fail to represent the actual local rainfall; (3) trends in the reanalyses may result from systematic changes throughout time in the availability of different types of observations. E.g., Hagemann *et al.* (2005) warns against trusting the trends in precipitation over land deduced from ERA40, because the precipitation bias differs from period to period, depending on the assimilation of satellite data. There may also be artificial trends in the re-analysis SLP; however, this is mostly a problem for regions with sparse observations, such as the Antarctic (Hines *et al.*, 2000). Thus, the poor agreement between the downscaled and estimated trends can be explained by the fact that the trends were not well defined and the reanalyses likely contain substantial errors. The fact that the downscaled results were not as sensitive to these arbitrary choices suggests that the part of the precipitation trends that can be associated with the large-scale structure is more robust, and hence gives a cautious reason for optimism in terms of downscaling climate scenarios.

The reason SLP-based PM was more skillful in reproducing past trends may further indicate that most of these observed recent trends actually were a consequence of circulation changes (Hanssen-Bauer and Førland, 2000, 1998) and the fact that SLP (which is strongly constrained by direct observations in the reanalyses) tends to have a

DOI: 10.1002/joc

higher quality than precipitation (which tend to be GCM predictions, given an assimilated atmospheric state).

The downscaling analysis can be regarded as a validation of both NCEP and ERA40 reanalyses with emphasis on a small selection of locations. The correlation and R^2 skill scores of DSMs calibrated with these two data sets were at most marginally different, suggesting that neither was superior to the other. The two data products exhibited similar abilities to reproduce past precipitation trends. ($r = 0.19$ and 0.16 for ERA40 and NCEP, respectively). All DSMs were most skillful in terms of trend reproduction for the winter and spring seasons and all performed poorly in summer and autumn. One reason for the precipitation models not performing well for summer and spring seasons may be because there are fewer cases with clear trends in these seasons. Furthermore, a great part of summer precipitation is associated with small-scale convective processes that are not related to large-scale SLPs in a straight-forward manner.

Testing the DSM predictions against 'true' values from the ECHAM5 results over independent sequences suggests that the DSMs do not suffer severely from non-stationary statistical relationships between large and small scales. Previous reports and present evaluations suggest that the large uncertainties in the precipitation products from the reanalyses preclude reliable tests of past trends. However, trend discrepancies may not affect the calibration of the DSM, which is based on de-trended data.

The results presented here were derived from the Nordic countries, but the difficulty related to the definition of trends in 50-year long precipitation data from the re-analysis is likely to be representative for other regions too. It is well known that much of the precipitation (orographic rain) along the west coast of Norway can be related to westerly winds (Henry, 1922), and hence SLP patterns (Nordli *et al*., 2005; Hanssen-Bauer *et al*., 2005), but SLP PMs may not be as skillful in other parts of the world, such as in the interior of continents. Thus, it is not given that the SLP PM yields the highest 'cor'- or r-scores for other regions. Furthermore, the test of stationarity was based on one GCM and the west coast of Norway, and the test results cannot be generalised to other GCMs or other locations.

ACKNOWLEDGEMENT

This work was done under the Norwegian Regional Climate Development under Global Warming (RegClim) programme, and was supported by the Norwegian Research Council (Contract NRC-No. 120 656/720) and the Norwegian Meteorological Institute. The analysis was carried out using the R data processing and analysis language (R Development Core Team, 2004), which is freely available over the Internet. We acknowledge the international modeling groups for providing their data for analysis, the PCMDI for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy.

APPENDIX

Downscaling details

Basics: principal component analysis

Temperature anomalies can be gridded and described mathematically in terms of a $M \times N$ matrix $\mathbf{X}_{\text{obs}}$. The observations made at time t can be represented by a vector $\vec{x}_t = [x_1, x_2, \ldots x_M]_t$, given by the columns of $\mathbf{X}_{\text{obs}}$, and one vector describes the spatial information at M locations for a given time t. If there are N observations made over a time interval, then the mathematical expression for the data can be written as $\mathbf{X}_{\text{obs}} = [\vec{x}_1, \vec{x}_2 \ldots \vec{x}_N]$. Likewise, the gridded temperature anomalies from the climate model scenarios can be represented as $\mathbf{X}_{\text{GCM}}$. Here the gridded observations (re-analysis) were de-trended by removing the best-fit linear temporal trend, since the presence of trends may bias the regression weights, and the trends in the reanalyses may not correspond to the trends in the predictand (station observations).

A singular value decomposition (SVD) was applied to the data matrix (SVD is described by Strang (1995) and Press *et al*. (1989)) in order to compute the right and left inverses of the data matrix $\mathbf{X}$ (here the subscript has been dropped).

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \tag{1}$$

In the above equation, the left inverse is a rectangular matrix, $\mathbf{U}$ and contains the right eigen-vectors in its columns, each of which describes an *eigen-pattern* (also referred to as 'modes') and is equivalent to an ordinary principal component analysis. Here, $\mathbf{U}$ has the dimensions $M \times K$, where K represents the total number of different modes. The SVD is equivalent to EOF analysis (Lorenz, 1963; North *et al*., 1982; Wilks, 1995; von Storch and Zwiers, 1999) when a geographical weighting is applied to the data so that the grid boxes representing smaller surface area carry proportionally less weight ($W_{ij} = \cos(\Phi_j)$, where $\cos(\Phi_j)$ is the latitude associated with the grid box). The right inverse, $\mathbf{V}$, is usually referred to as the *principal components* in the geophysical literature, and is a matrix that describes the time evolution of the various eigen-patterns. It has $N \times K$ dimensions, and the columns hold a set of weights used to determine how much each mode contributes at a given time. The principal components were divided into 2 subsections: $\mathbf{V}^T = [\mathbf{V}^T_{\text{dep}}, \mathbf{V}^T_{\text{ind}}]$ for subsequent analysis, where one subsection contained dependent data for calibration and the other contained independent data for validation. The symbol $\mathbf{\Lambda}$ in equation 1 is a diagonal matrix with the

DOI: 10.1002/joc

eigen-values given in descending order along its diagonal. It is common to discard the higher order modes in order to simplify regressional analysis and filter away noise.

A geographical weighting was used before the SVD and an inverse scaling function was applied to the spatial patterns **U** obtained from the SVD. The data were sub-sampled in accordance with North *et al*. (1982) to reduce any effects caused by autocorrelation. The time interval used for sub-sampling was determined according to the lag at which the serial correlation approached zero. The estimate for **V** is then computed by projecting the entire data set onto the eigen-patterns computed using the sub-sampled data: $\mathbf{V}^T = \mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{X}$.

The mixed-field predictor: mixFields

The mixed-field predictor consists of two different physical elements, precipitation and SLP, combined to form a new data matrix with increased size. If the precipitation is represented by an $M_1 \times N$ matrix and SLP by an $M_2 \times N$, then the mixed-field is given by an $(M_1 + M_2) \times N$ matrix, of which the columns are $\vec{x}_t = [y_1, y_2, \ldots y_{M_1}, z_1, z_2, \ldots, z_{M_2}]_t$, y are standardised values of the monthly precipitation and z the standardised values of SLP. Standardisation is used to give the two fields equal weight, and the mean standard deviation for the whole respective field is used: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x-\overline{x})^2}$, where $n = M_1 \times N$ for precipitation and $n = M_2 \times N$ for SLP. Here, $\overline{x}$ is the temporal-spatial mean value. In this case, $M_1 = M_2$ as the precipitation and SLP were given on the same grid.

The downscaling procedure: DS

In the downscaling analysis implemented by the **DS** function in the `clim.pact` package, the subsection of the PCs containing the dependent data was used for a model calibration involving a stepwise multiple regression (Wilks, 1995). The empirical models were seasonally stratified, with one respective model for each calendar month. A backward–forward stepwise screening using the Akaike Information Criterion (AIC) (Wilks, 1995, p.194-198) was employed to determine whether each pattern should be included or not. The regression analysis can be described by

$$\hat{Y} = \Psi\mathbf{V}_{\text{dep}}^T \tag{2}$$

where $\hat{Y}$ is the solution of equation 2, which gives the lowest root-mean-square-error over the calibration set, and the matrix Ψ is the statistical model that can be used for prediction if $\mathbf{V}_{\text{dep}}^T$ in equation 2 is replaced by $\mathbf{V}_{\text{ind}}^T$. The data in $\mathbf{V}_{\text{dep}}^T$ and the observations Y were de-trended prior to the Principal Component Analysis (PCA) and the regression analysis.

The split-merge strategy: mergeStation

The split-merge strategy consists of dividing the data into two parts according to their chronology. The first half (X_1) represents the earliest observations whereas the other half (X_2) consists of the most recent data. One set of DSMs are calibrated with X_1, henceforth referred to as Ψ_1, and are subsequently used together with the independent data to make a prediction for the second period: thus $[y_1 X_1] \rightarrow \Psi_1$, $\hat{y}_2 = \Psi_1 X_2$. Conversely, $\hat{y}_1 = \Psi_2 X_1$. There are then two independent predictions, $\hat{y}_1$ and $\hat{y}_2$, which are merged in order to get one long series to which the trend analysis is applied. To ensure that the long-term trends are not affected by artificially different constant levels in the two series, the last series was adjusted so that it had the same mean value over the 1976–1984 interval: $\hat{y}_2 = \hat{y}_2 - \frac{1}{9}\left[\sum_{1976}^{1984}(\hat{y}_2) + \sum_{1976}^{1984}(\hat{y}_1)\right]$.

REFERENCES

Benestad RE. 2000. Analysis of gridded sea level pressure and 2-meter temperature for 1873–1998 based on UEA and NCEP re-analysis II. KLIMA 03/00. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad RE. 2001. A comparison between two empirical downscaling strategies. *International Journal of Climatology* **21**: 1645–1668, DOI 10.1002/joc.703.

Benestad RE. 2002. Empirically downscaled multi-model ensemble temperature and precipitation scenarios for Norway. *Journal of Climate* **15**: 3008–3027.

Benestad RE. 2003. What can present climate models tell us about climate change? *Climatic Change* **59**: 311–332.

Benestad RE. 2004a. Empirical-statistical downscaling in climate modeling. *EOS* **85**(42): 417.

Benestad RE. 2004b. *Empirically Downscaled SRES-based Climate Scenarios for Norway, Climate 08*. The Norwegian Meteorological Institute: Oslo, Norway; www.met.no.

Benestad RE, Melsom A. 2002. Is there a link between the unusually wet autumns in southeastern Norway and SST anomalies? *Climate Research* **23**: 67–79.

Busuioc A, Chen D, Hellström C. 2001. Performance of statistical downscaling models in GCM validation and regional climate change estimates: application for Swedish precipitation. *International Journal of Climatology* **21**: 557–578.

Chen D, Achberger C, Räisänen J, Hellström C. 2005. Using statistical downscaling to quantify the GCM-related uncertainty in regional climate change scenarios: a case study of Swedish precipitation. *Advances in Atmospheric Sciences* **23**(1): 54–60.

Hagemann S, Arpe K, Bengtsson L. 2005. *Validation of the Hydrological Cycle of ERA-40*, *ERA-40 Project Report Series 24*. ECMWF: England; www.ecmwf.int, Reading, RG2 9AX.

Hanssen-Bauer I, Førland EJ. 1998. Long-term trends in precipitation and temperature in the Norwegian Arctic: can they be explained by changes in the atmospheric circulation patterns? *Climate Research* **10**: 143–153.

Hanssen-Bauer I, Førland E. 2000. Temperature and Precipitation variations in Norway 1900–1994 and their links to atmospheric circulation. *International Journal of Climatology* **20**: 1693–1708.

Hanssen-Bauer I, Førland EJ, Haugen JE, Tveito OE. 2003. Temperature and precipitation scenarios for Norway: comparison of results from dynamical and empirical downscaling. *Climate Research* **25**: 15–27.

Hanssen-Bauer I, Achberger C, Benestad RE, Chen D, Førland EJ. 2005. Statistical downscaling of climate scenarios over Scandinavia: a review. *Climate Research* **29**: 255–268.

Hellström C. 2003. Regional precipitation in Sweden in relation to large-scale climate, PhD thesis, Göteborg University.

Hellström C, Chen D, Achberger C, Räisanen J. 2001. Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. *Climate Research* **19**: 45–55.

Henry AJ. 1922. J. Bjerknes and H. Solberg on meteorological conditions for the formation of rain. *Monthly Weather Review*: **50**(8): 402–404.

Hines KM, Bromwich DH, Marshall GJ. 2000. Artificial surface pressure trends in the NCEP-NCAR reanalysis over the Southern Ocean and Antarctica. *Journal of Climate* **13**: 3940–3952.

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA. 2001. Contribution of Working Group I to the Third Assessment Report of IPCC. International Panel on Climate Change. *Climate Change 2001: The Scientific Basis*. Cambridge University Press: Cambridge, UK; (Available from www.ipcc.ch).

Imbert A, Benestad RE. 2005. An improvement of analog model strategy for more reliable local climate change scenarios. *Theoretical and Applied Climatology* **82**: 245–255, DOI: 10.1007/s00704-005-0133-4).

IPCC. 1995. *The Second Assessment Report. Technical Summary*. WMO & UNEP. Cambridge University Press: Cambridge, UK; http://www.ipcc.ch/pub/sa(E).pdf.

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Wollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77**: 437–471.

Linderson M-L, Achberger C, Chen D. 2004. Statistical downscaling and scenario construction of precipitation in Scania, southern Sweden. *Nordic Hydrology* **35**: 261–278.

Lorenz E. 1963. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **20**: 130–141.

Nordli O, Lie O, Nesje A, Benestad RE. 2005. Glacier mass balance in southern Norway modelled by circulation indices and spring-summer temperatures ad 1781–2000. *Geografiska Annaler Series A-Physical Geography* **87A**(3): 431–445.

North GR, Bell TL, Cahalan RF. 1982. Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review* **110**: 699–706.

Oberhuber JM. 1993. Simulation of the Atlantic circulation with a coupled sea ice-mixed layer isopycnal general circulation model. Part 1: model description. *Journal of Physical Oceanography* **22**: 808–829.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT. 1989. *Numerical Recipes in Pascal*, Cambridge University Press, New York, USA; pp 759.

R Development Core Team. 2004. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna; ISBN 3-900051-07-0.

Roeckner E, Arpe K, Bengtsson L, Dümenil L, Esch M, Kirk E, Lunkeit F, Ponater M, Rockel B, Sausen B, Schlese U, Schubert S, Windelband M. 1992. Simulation of present-day climate with the ECHAM model: impact of model physics and resolution. Technical Report 93. Max Planck-Institute für Meteorologie: Hamburg.

Simmons AJ, Gibson JK. 2000. *The ERA-40 Project Plan*, *ERA-40 Project Report Series 1*. ECMWF: Reading, UK; www.ecmwf.int.

Strang G. 1995. Linear Algebra and its Application.

Tuomenvirta H, Drebs A, Førland E, Tveito OE, Alexandersson H, Laursen EV, Jónsson T. 2001. Nordklim data set 1.0. KLIMA 08/01. met.no, P.O.Box 43 Blindern, N-0313 Oslo, Norway, (www.met.no).

von Storch H, Zwiers FW. 1999. *Statistical Analysis in Climate Research*. The Norwegian Meteorological Institute: Oslo, Norway.

von Storch H, Hewitso nB, Mearns L. 2000. Review of empirical downscaling techniques. In *Regional Climate Development Under Global Warming*, *General Technical Report 4*, Iversen T, Høiskar BAK (eds). http://regclim.met.no/rapport_4/presentation02/presentation02.htm. NILU: Kjeller, Norway.

Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences*, Academic Press: San Diego, CA; 464–467.