Theor. Appl. Climatol. 87, 123–137 (2007)
DOI 10.1007/s00704-005-0223-3

Theoretical and Applied Climatology
Printed in Austria

[1] Air and Water Science, Department of Earth Sciences, Uppsala University, Villavägen, Uppsala, Sweden
[2] Department of Geosciences, University of Oslo, Norway

Seasonality properties of four statistical-downscaling methods in central Sweden

F. Wetterhall[1], **S. Halldin**[1], and **C.-Y. Xu**[2]

With 7 Figures

Received December 8, 2004; revised June 20, 2005; accepted July 1, 2005
Published online June 14, 2006

Summary

Daily precipitation in northern Europe has different statistical properties depending on season. In this study, four statistical downscaling methods were evaluated in terms of their ability to capture statistical properties of daily precipitation in different seasons. Two of the methods were analogue downscaling methods; one using principal component analysis (PCA) and one using gradients in the pressure field (Teweles-Wobus scores, TWS) to select the analogues in the predictor field. The other two methods were conditional-probability methods; one using classification of weather patterns (MOFRBC) and the other using a regression method conditioning a stochastic weather generator (SDSM). The two analogue methods were used as benchmark methods. The study was performed on seven precipitation stations in south-central Sweden and the large-scale predictor was gridded mean-sea-level pressure over Northern Europe. The four methods were trained and calibrated on 25 years of data (1961–1978, 1994–2000) and validated on 15 years (1979–1993). Temporal and spatial limitations were imposed on the methods to find the optimum predictor settings for the downscaling. The quality measures used for evaluating the downscaling methods were the residuals of a number of key statistical properties, and the ranked probability scores (RPS) for precipitation and maximum length of dry and wet spells. The results showed that (1) the MOFRBC and SDSM outperformed the other methods for the RPS, (2) the statistical properties for the analogue methods were better during winter and autumn; for SDSM and TWS during spring; and for MOFRBC during summer, (3) larger predictor areas were needed for summer and autumn precipitation than winter and spring, and (4) no method could well capture the difference between dry and wet summers.

1. Introduction

General circulation models (GCMs) are important tools in evaluating the effects of climate change as caused by the increase of radiatively-active gases on a global scale. Since catchment-based simulation of runoff requires input on a much higher resolution than what can be provided by these global models, precipitation-downscaling tools are required to study the effects of a global change on local and regional scales. Runoff in central Sweden has a clear seasonal variation with lower values during winter and summer months, and higher values during spring flood and early autumn (Xu et al., 1996). In order to model the seasonality in daily runoff in a perturbed climate, it is important to correctly capture the seasonal cycle of the daily precipitation in statistical downscaling studies. The seasonal variation of both the mean and the variability of precipitation also have an important impact on the length of the vegetation period (Barrow and Semenov, 1995).

Earlier studies on statistical precipitation downscaling have been performed both on a single season, for example daily winter precipitation (Zorita and von Storch, 1999; Biau et al., 1999) and autumn precipitation (Obled et al., 2002), and on two or more seasons (Beckmann and Buishand, 2002; Hay et al., 1991; Stehlik and Bardossy, 2002;

Wilby et al., 1998). Methods of statistical precipitation downscaling can be categorised in many ways and we choose to classify them into three main groups: Analogue or resampling, regression, and weather classification methods. For a more extensive description of the methods see Wilby and Wigley (1997) and Xu (1999). The methods that include two or more seasons address seasonality in different ways. The typical approach in regression models is to generate different regression parameters for different seasons or for different months. The methods of weather classification use seasonally-dependent precipitation generators conditioned on the classification. If a method is constructed with different parameter values for different seasons, the imposed seasonality has to be valid for a future perturbed climate (Kilsby et al., 1998). Earlier studies have also noted that circulation patterns influence the precipitation amount and occurrence differently depending on the season (Zorita et al., 1995).

The aim of the study was to evaluate the seasonality properties of four different methods in downscaling daily precipitations from the large-scale mean-sea-level-pressure (MSLP) field and compare method downscaling skills within each season. Two analogue methods (i.e. principal-component analysis (PCA) and Teweles-Wobus Scores (TWS)), one weather-pattern method (MOFRBC), and one regression method (SDSM) with a weather generator were selected for this study.

Since precipitation is an inherently stochastic and non-linear process, the evaluation focused on the statistical properties (i.e. distributions and key variables rather than on the precipitation amounts on single days). The seasonality properties of the downscaling methods were analysed both with respect to the inherent methodological features of each method and spatial and temporal calibration of the predictor–predictand relationships.

2. Study region and data

2.1 Predictands

The predictand (i.e. the variable to be downscaled) was daily precipitation. Daily data from 19610101–20001231 were purchased from the Swedish Meteorological and Hydrological Institute (SMHI) for 10 stations close proximal to Uppsala in south-central Sweden (Fig. 1). The stations were selected within the southern NOPEX region (Halldin et al., 1999), approximately at 60° N latitude and 18° W longitude, in order to support other climate-related research carried out for this region. We retained data from 7 stations (Table 1) for which the data sets were reasonably complete. Slightly less than 6% of the data from one or more stations was missing for the entire period; whereas the time series was 100% complete for the period 1974–1997. Since precipitation occurrence and amounts are stochastic by nature, gap-filling would not improve results. Therefore, we did not pre-process data this way. The data we received were original measurements. SMHI also supplied correction factors to

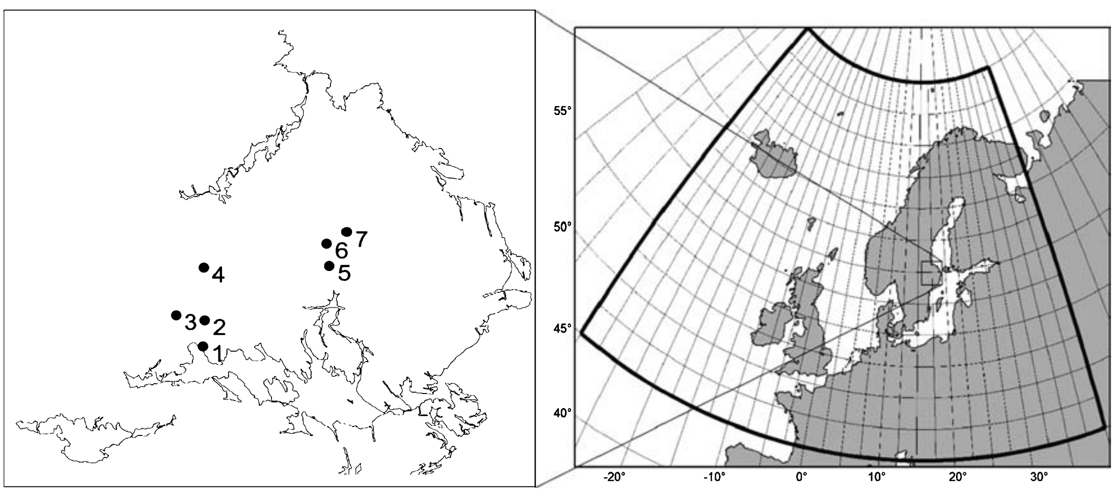

Fig. 1. Location of precipitation stations in the target area and extent of the $2.5° \times 2.5°$ mean-sea-level-pressure predictor dataset. Modified from Wetterhall et al. (2005)

Table 1. The mean annual precipitation (1961–1990) and the coordinates of the precipitation stations used in the study

No.	Station	Latitude	Longitude	Prec. (mm)
1	Västerås-Hässlö	59°35′51″	16°37′57″	561
2	Sundby	59°41′46″	16°39′38″	659
3	Skultuna	59°42′50″	16°26′10″	656
4	Sala	59°54′16″	16°39′38″	637
5	Uppsala airport	59°53′43″	17°35′36″	599
6	Drälinge	59°59′32″	17°34′25″	615
7	Vattholma	59°1′44″	17°43′27″	657

account for wind loss, adhesion to, and evaporation from measurement vessels at each station (Eriksson, 1983). We used the corrected data in this study.

2.2 Predictors

The selection of an appropriate predictor, or characteristics from the large-scale atmospheric circulation, is one of the most important steps in a downscaling exercise. Three main factors constrain the choice of predictors. Data should be (1) reliably simulated by GCMs, (2) readily available from archives of GCM output, and (3) strongly correlated with the surface variables of interest (Wilby et al., 1999). Daily MSLP fulfils the two first criteria and has been widely used in downscaling of precipitation (e.g. Barrow and Semenov, 1995; Zorita et al., 1995; Kilsby et al., 1998; Conway and Jones, 1998; Stehlik and Bardossy, 2003) and has a documented correlation with precipitation (e.g. Wetterhall et al., 2005). In this study, only MSLP was used in order to compare the different methods in an objective manner. The predictors covering an area of 45°–75° N, 40° W–40° E were downloaded from the NCEP/NCAR reanalysis project (Kalnay et al., 1996; http://dss.ucar.edu/pub/reanalysis/). The geographical extent was chosen to include all areas with noticeable influence on the circulation patterns that govern weather in Scandinavia (Hanssen-Bauer and Førland, 2000). The dataset had a grid resolution of $2.5° \times 2.5°$ longitude–latitude.

3. Methods

Four methods to downscale precipitation were used in this study (i.e. two benchmark methods using historical analogues, one classification method using weather patterns, and one regression method). For the sake of completeness, a brief description of the methods is presented in the following sub-sections. See the references in Table 2 for a more detailed description of each method.

3.1 Analogue methods

Analogue methods use historical data sets of predictor and predictand to model the predictand. The methods search an historical database of predictors and sample the predictor from this dataset that best resembles the predictor on the day on which to downscale the predictand (target day) (Zorita and von Storch, 1999; Obled et al., 2002).

Table 2. Statistical-downscaling methods used in the study

Notation	Description	References
PCA	Analogue method that uses principal component analysis to select the most suitable analogue from an historical dataset of mean-sea-level pressure	Cubasch et al. (1996), Zorita and von Storch (1999)
TWS	Analogue method that compares the N–S and E–W gradients in the mean-sea-level-pressure field to select the most suitable analogue from an historical dataset	Teweles and Wobus (1954), Obled et al. (2002)
MOFRBC	A weather-pattern-classification method using fuzzy rules. The patterns can be objectively or subjectively derived. The method has been used in European mainland and Great Britain	Bardossy et al. (1995), Stehlik and Bardossy (2002)
SDSM	Statistical-downscaling model (SDSM) is a multivariate regression model that uses a weather generator to model a predictand from derived regression equations. The model has been used in Great Britain and North America	Wilby et al. (2002)

The analogue of the target day is selected by analysing circulation patterns of the predictor field. In this application, the entire time series was used to select the analogue; with the exception that an analogue could not be selected from the same year as the target pattern. The predictor, in this case the MSLP pattern, was modelled in the calibration and validation periods by choosing the analogue to the MSLP patterns in the training period. The generated precipitation was the precipitation resampled from the selected occurrence in the training period. The benefits of the methods are that they can easily be applied to predictands that are not normally distributed and the spatial correlation of the predictand is preserved. The primary drawback is that only events that have occurred in the past can be modelled; possibly limiting the validity in a perturbed climate (Zorita and von Storch, 1999). The methods provide no physical interpretation of the relationship between predictor and predictand (Zorita et al., 1995), but the introduced areal and temporal restrictions on the predictor should be physically valid. The methods were used in this study as benchmarks against which the two more sophisticated methods were compared.

The two techniques used in selecting the analogues from the MSLP were PCA and TWS. The PCA, also known as Empirical Orthogonal Functions (EOFs) in meteorological applications, basically determines the internal relationship of the anomalies of the predictors and relates them to the predictand (Huth and Kysely, 2000). The TWS method compares the gradients instead of the anomalies at each point in the predictor field. This method was first used as a quality measure of geopotential-height forecasts (Teweles and Wobus, 1954), but has been used in downscaling studies in recent years as a predictor of precipitation in flood forecasting (Obled et al., 2002). The two methods have been evaluated over the study area and it was found that both methods reproduced daily and monthly precipitation characteristics well enough to be considered as benchmarks in comparison studies (Wetterhall et al., 2005).

3.2 Weather classification method

MOFRBC is a conditional-probability method based on an objective weather-pattern classification (Bardossy and Plate, 1992; Bardossy et al., 1995) that has been used for downscaling in the British Isles, Central Europe, Germany and Greece (Stehlik and Bardossy, 2002, 2003). The method is based on fuzzy rules which provide a statistical method for classifying an event into a specific predetermined classification using logical and probabilistic statements (Bardossy et al., 1995). The term fuzzy refers to the logical expression, or rule, dependant on which class the weather pattern belongs to and is associated with a certain degree of fulfilment (DOF). An event is classified by assigning it to the weather pattern which has the highest DOF. The scheme can use any type of predictor (MSLP, GPH) and local weather variable as predictand (precipitation, temperature, runoff, et cetera).

The predictand is stochastically modelled conditioned on an intermediate predictor, the weather-classification patterns. These patterns can either be built on existing subjective classification schemes or objectively derived patterns. MOFRBC optimises the weather patterns to maximise variability in the predictand to give weather situations of different character. In this study, the 12 most common patterns from European Grosswetterlagen (Baur et al., 1944) were adapted to the study region and optimised to give precipitation patterns of different types i.e. very dry or very wet conditions. The measure of a patterns condition is defined by its wetness:

$$Wet(i) = \frac{\frac{1}{P}\sum_{t=1}^{T} p(t_i)}{\frac{\sum t_i}{T}} - 1 \tag{1}$$

where p is the total amount of precipitation for the day t classified in pattern i, P is total amount of precipitation for all the T classified days. In order to achieve negative values for dry patterns, 1 is subtracted from the primary term. Days that were not classified in any of the 12 patterns were allocated to a residual group.

The classifications were optimised and evaluated according to criteria depending on precipitation occurrence (I_1) and precipitation amount (I_2).

$$I_1 = \frac{1}{T}\sum_{t=1}^{T} \sqrt{\left(p(\mathrm{CP}(t)) - \bar{p}\right)^2} \tag{2}$$

where T is the number of classified days, $p(\mathrm{CP}(t))$ is the probability of precipitation on day t with classification CP, and $\bar{p}$ is the probability of precipitation for all days.

$$I_2 = \frac{1}{T}\sum_{t=1}^{T}\left|\ln\left(\frac{z(\mathrm{CP}(t))}{\bar{z}}\right)\right| \tag{3}$$

where z is the mean precipitation amount on day t with classification CP, T is the number of classified days. The patterns are hybrids between subjectively and objectively classified patterns, having the advantage of the physically meaningful Grosswetterlagen, but modified to capture differences in precipitation regimes. The trade-off is that the patterns are not totally conditioned on precipitation; also a wet pattern has dry days. Enke et al. (2005) have proposed a subjective classification scheme that uses whatever predictor set that provides the optimum result and classifies circulation patterns according to predictand intervals. This approach is appealing concerning the precipitation statistics for each pattern, but it also makes it even more important to make sure that the patterns are physically reasonable and not just mathematical artefacts.

3.3 Regression method

The Statistical-Downscaling Model (SDSM) uses a multivariate linear regression method to derive a statistical relationship between predictor and predictand and has been applied to a number of catchments in Great Britain and North America (Wilby et al., 2002). Predictors can be any large-scale atmospheric variables that have a correlation with the predictand. The method includes built-in transform functions in order to obtain secondary data series of the predictand and/or the predictor that have stronger correlations than the original data series (Wilby et al., 2002). The precipitation is then modelled through a weather generator conditioned on the predictor variables. The problem of over-fitting caused by collinearity (i.e. too many variables and parameters in the model that are highly correlated) has to be addressed when building a regression model (Wilby and Wigley, 2000; Enke et al., 2005). Over-fitting problems tend to produce a calibration result that is too optimistic and have the tendency to perform poorly when presented with new observations or instances (i.e. they do not generalise well to the prediction of "new" cases). One of the methods to handle this problem is reducing the number of variables in the model or increasing the size of the sample. The predictors with highest Spearman-ranked correlation with precipitation were all highly inter-correlated (>0.8) so the best model should be built using only the predictor with the highest explained variance in the predictand. However, using only one variable resulted in under-fitting the problem, meaning that the model is not sufficiently complex to fully detect the signal in a complicated data set. In our case, the duration of wet and dry spells were not captured by the model using only one variable. Under-fitting produces excessive bias in the outputs, whereas over-fitting produces excessive variance. To compromise, the chosen predictors were extended to six grid points of MSLP and two grid points of 1-day-lagged MSLP using the highest correlation with the predictand. This approach introduced some collinearity in the predictors and over-fitted the stochastic weather generator to some extent, but since the duration of dry and wet spells were improved and no significant change could be detected in the other objective functions, this model setup was preferred.

4. Method evaluation

The comparison study was carried out by applying the same training data set to all methods. The 40-year precipitation-data series was broken up into a calibration period and a validation period. The methods were calibrated on the training period 1961 to 1978 and 1994 to 2000. Method validation involved period 1979 to 1993. These time periods were selected to agree with the periods used in the European project for inter-comparison of statistical downscaling methods, STARDEX (2001). The validation period, 1979 to 1993, was selected to agree with the ERA-15 reanalysis period.

4.1 Objective functions

The method evaluation used objective functions representing 3 precipitation-distribution properties and 4 key precipitation variables (Table 3) suggested by the STARDEX (2001) project. The STARDEX variables were selected to capture extreme probabilities in precipitation series rather than monthly or yearly totals.

Table 3. Objective functions and evaluation distributions/variables. Evaluation variables are selected from the core indices suggested by the STARDEX (2001) project

Objective function	Evaluation distribution/variable
Ranked-probability score	Precipitation distribution (RPS_{prec}) Dry-spell-length distribution (RPS_{dry}) Wet-spell-length distribution (RPS_{wet})
Residual function	Average-wet-day amount (RF_{amount}; mm/day) 90th-percentile-of-rain-day amounts ($RF_{90\%}$; mm/day) Greatest 5-day total rainfall (RF_{max}; mm) Maximum length of dry spell (RF_{dry}; days)

The first objective function was the ranked-probability score (*RPS*), developed to evaluate the probability distribution of an ensemble of forecasts compared to the observed values. In this case RPS was calculated for daily precipitation and seasonal maximum duration of dry and wet spells (Epstein, 1969; Murphy, 1971; Obled et al., 2002). To define *RPS* we consider a random variable X with $K > 2$ thresholds $x_1 < x_2 < \cdots < x_k$ that define the events $A_k = \{X \leq x_k\}$ for $k = 1,2,\ldots K$ with the forecast probabilities $(\hat{p}_k, \hat{p}_k, \ldots, \hat{p}_k)$. The binary indicator variable for the k^{th} event is denoted o_k and defined as $o_k = 1$ if A_k occurs and 0 otherwise. *RPS* is then

$$RPS = 100\frac{1}{N}\sum_{n=1}^{N} B_k \tag{4}$$

$$B_k = \frac{1}{K}\sum_{k=1}^{K}(\hat{p}_k - o_k)^2 \tag{5}$$

where N is the number of forecasts. The factor 100 in Eq. (4) is used to scale the *RPS* values to a range that is convenient to use. A value close to 0 denotes a good simulation. *RPS* has a subjective component, i.e. the choice of the number of classes and the limits of each class. When used for precipitation quantity, 8 classes were established: 0 (no rain), 0–1, 1–3, 3–5, 5–10, 10–20, 20–50, and more than 50 mm rain per day. When used for maximum dry-spell length, the established classes were <7, 7–9, 10–12, 13–16, 17–20, and >20 days. For maximum wet-spell lengths the established classes were <5, 5–7, 8–10, 11–14, 14–20 and >20 days. The classes were established to reflect the distribution in the observed data and were the same for all stations. In the results *RPS* was averaged over all stations.

The second objective function calculates and compares the difference of the statistical properties between observed and calibrated precipitation series. This function is henceforth referred to as the residual function (*RF*):

$$RF(y_j) = \frac{1}{N}\sum_{n=1}^{N}[y_{obs}(n) - y_{sim}(n,j)]^2 \tag{6}$$

where y is the statistical property of the j^{th} simulated and observed precipitation series respectively and N is the number of stations. A value close to zero for the residual function implies good downscaling. The *RF* value and *RPS* were consistently used during the calibration to optimise the settings for each method. The standard deviation for the observed values was calculated from yearly means over the validation period.

4.2 Spatial coverage

The first step in the downscaling was to decide on the optimal predictor grid-point area. The procedure differed somewhat between methods. The procedure for the analogue methods was to exclude grid points, one by one, and graphically evaluate downscaling for each objective function. The spatial variability of the objective function, *OF*, is expressed as

$$OF(y_i) = \sum_{k=1}^{K} RF(y_k), \quad k = 1, 2, \ldots, K, \ k \neq i \tag{7}$$

where i denotes the simulation with the excluded grid point, and K are the included grid points. Area windows of varying size and orientation over the study area were constructed following the result of this analysis. These windows ranged from a local window covering only the area above the stations to one including the entire predictor data set. The area windows were used in the optimisation of the two analogue methods and the MOFRBC method. Twelve large-scale circulation patterns derived from the

European Grosswetterlagen, describing cyclonic, anticyclonic, N–S and E–W circulation, were used as input classification for the latter method.

For the SDSM, the Spearman-rank-correlation coefficient for the predictor, 1-day-lagged predictor data, and predictand data were analysed. We used the 6 grid points of MSLP data combined with the 2 one-day-lagged MSLP data that had the highest correlation with precipitation in the study (Fig. 6).

4.3 Evaluation of seasonality

Precipitation distribution in Sweden is highly affected by elevation (Johansson and Chen, 2003) and precipitation quantities show distinct seasonal cycles having maximum precipitation in August and a minimum during winter. Variability is largest during summer months. The construction, calibration and validation of each method were performed on four seasons, winter (DJF), spring (MAM), summer (JJA) and autumn (SON), in order to achieve optimum seasonality performance of the methods. The approach in dealing with seasonality differed between methods. Seasonality was considered implicitly in the analogue methods by using a time window such that an analogue could only be selected from training periods in the same time of year as the target day. For example, if the target day was 1 June and the optimum time window ± 17 days, then analogues were selected from the range 14 May to 18 June in the data set. The optimum time-window size was selected by simulating precipitation with all possible combinations of spatial and temporal windows.

The MORFBC includes seasonality implicitly through the use of seasonal variation of the precipitation variance (Stehlik and Bardossy, 2003). The seasonality was investigated by separately analysing the precipitation patterns provided by each classification for different seasons. Seasonality, in SDSM, was explicitly expressed by comparing the Spearman-rank-correlation coefficient between predictor and predictand and then selecting the optimum predictor set for each season. The objective functions were then investigated for each season and the model was built accordingly. The resulting parameter-value sets for the model were, therefore, different for each season.

The methods were also evaluated for the five wettest and five driest summers in order to compare how the methods processed different climate situations within the study area and time period. The selection of wet and dry summers was based on the total precipitation amount for the specific season. It is noticeable that all the wettest summers occurred during the validation period and all the driest summers during the calibration period. All results were averaged from an ensemble of 20 downscaled simulations in order to derive results that were representative of the method. This was straightforward with MOFRBC and SDSM because they produce stochastic time series. In contrast, this is problematic with the analogue methods because they are deterministic. Ensembles of analogue downscaled time series were created by including simulations with time and area windows that were close to optimal. The ensemble included simulations with area window ± 1 and time window ± 3 resulting in 21 simulations. An ensemble of 20 was produced by excluding one randomly selected simulation.

5. Results

5.1 Main circulation modes

The four leading principal components (PCs) of the PCA method explained 83% of the winter variance; 73% of the spring and autumn variance; and 74% of the summer variance of the MSLP-anomaly field. The anomaly field is basically the same over the year (Fig. 2).

The MSLP-anomaly patterns obtained from classification after the optimisation procedure with the MOFRBC method indicated similarities between different patterns (Fig. 3). The patterns in Fig. 3 exemplify the summer classification. The most frequently occurring wet pattern was CP_{11}. However, this pattern was not the wettest during summer (Fig. 4). Rather, the wettest patterns were associated with CP_6 and CP_{12} which involve a strong cyclonic anomaly centre over the study area and corresponding anticyclonic activity around Iceland. The dry CP_3 closely resembled a mirror image of the wet CP_6; the latter with a cyclonic anomaly over the study area. The frequency and wetness index varied depending on season (Fig. 4).

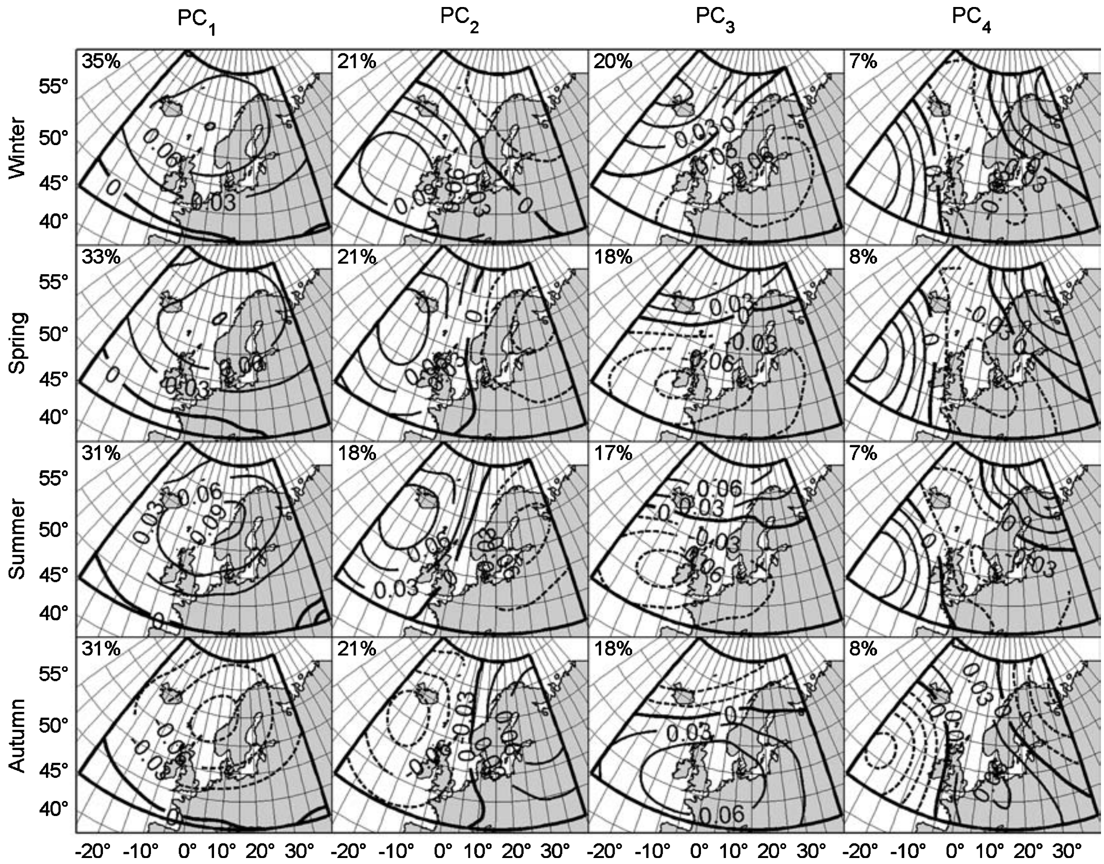

Fig. 2. The four leading principal components (PCs) of the daily 1961–1990 mean-sea-level-pressure field during winter (DJF), spring (MAM), summer (JJA) and autumn (SON). The percentage is explained variance of each PC

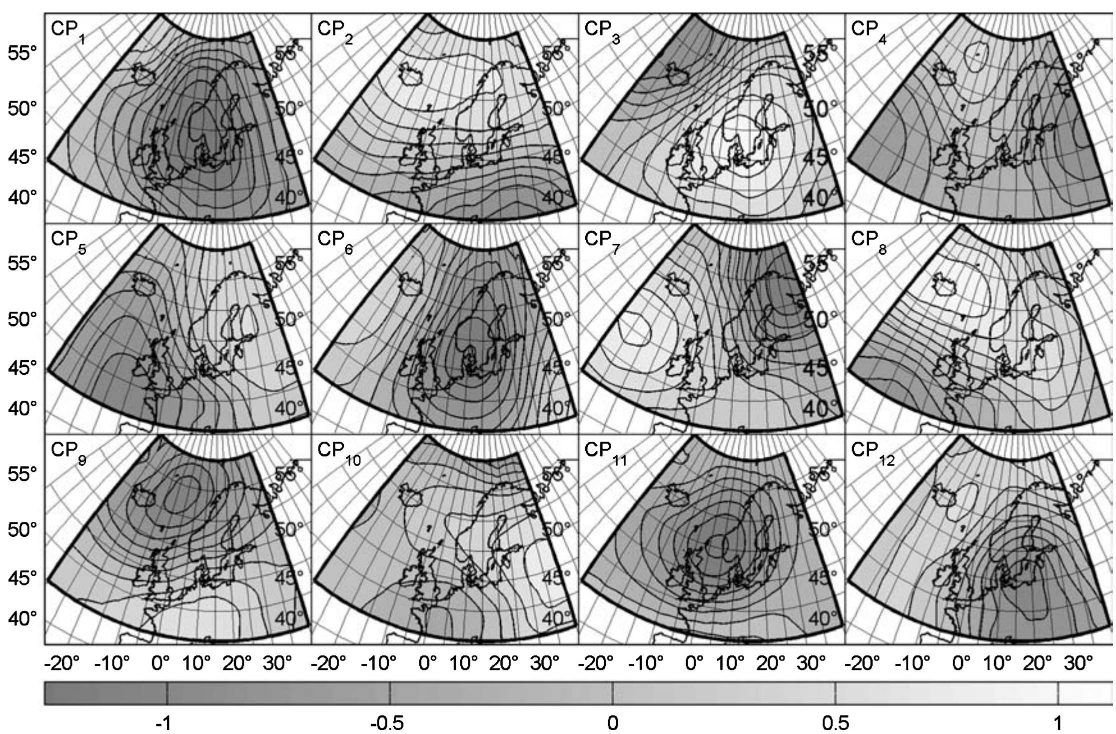

Fig. 3. Composite maps of 1961–1978 mean-sea-level-pressure-anomaly circulation patterns (CPs) for summer according to the multiobjective-fuzzy-rule-based classification method

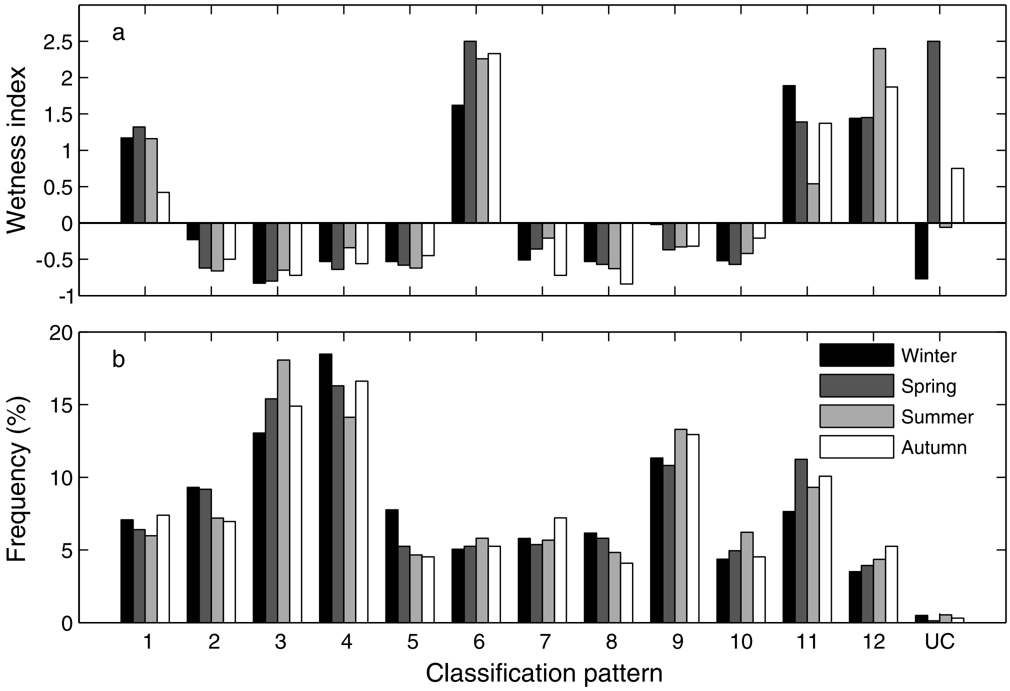

Fig. 4. Wetness index and frequency for the 12 different circulation patterns (CPs) of the MOFRBC method. UC represents unclassified days

5.2 Spatial and temporal calibration

The optimum-area windows for the analogue methods varied highly in size depending on season (Fig. 5). Both TWS and PCA methods had larger optimum areal windows for summer season than winter and spring. In addition, TWS also had a larger autumn window. The time windows for the TWS and PCA methods were 5 and 20 days for winter, 7 and 15 days for spring, 15 and 35 days

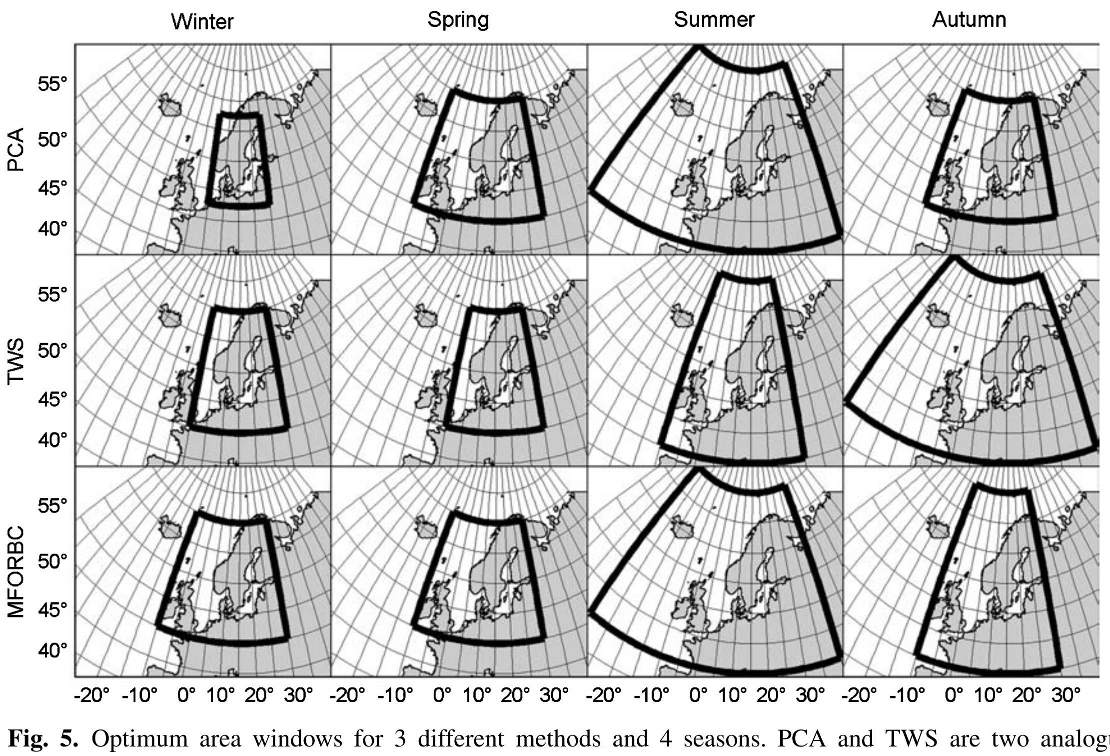

Fig. 5. Optimum area windows for 3 different methods and 4 seasons. PCA and TWS are two analogue methods, and MOFRBC is a weather-classification method

for summer, and 13 and 15 days for autumn, respectively.

The optimum areal windows for the MOFRBC were extracted with the same procedure used with the analogue methods, but the evaluation was completed with the selection criteria I_1 and I_2. This resulted in the largest areal window in the summer.

The screening of the MSLP grid in the SDSM method provided different results depending on season. Winter precipitation had its maximum, non-lagged, negative correlation located in the south–west of the study area (Fig. 6). The grid points directly above the stations were not included in the regression for winter because these grid points were not sufficiently correlated with precipitation. The maximum negative correlation was concentrated over the study area during the summer. A weak negative correlation field was associated with the area around Iceland. The one-day-lagged grid points with the highest correlation were consistently east of the non-lagged points. There was no overlap between these points except for one point in Västerås during summer. The correlation field for spring was similar to the winter field and the autumn correlation field resembled the summer field.

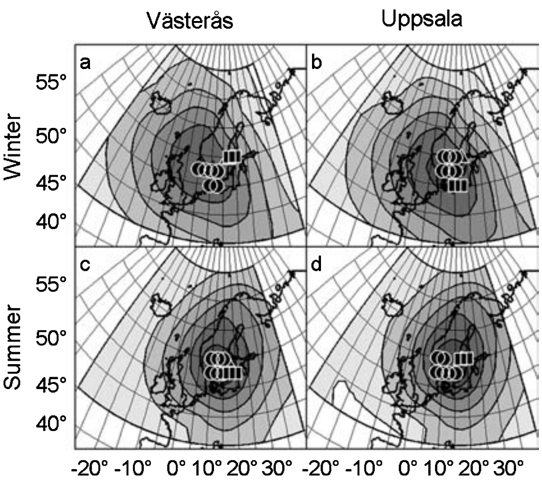

Fig. 6. Spearman-rank-correlation coefficient between instantaneous (circles) and 1-day-delayed (squares) mean-sea-level pressure (MSLP) and winter/summer precipitation in Västerås and Uppsala. Circles and squares show grid points used in the regression. The triangles give the station locations

5.3 Intra-annual variability

All methods simulated the intra-annual variation well, but the general result was that all methods under-estimated summer precipitation (Fig. 7). The SDSM method best captured monthly summer precipitation. Winter precipitation was generally best downscaled.

5.4 Measures of precipitation-downscaling skill

The MOFRBC method performed best concerning downscaling the RPS_{prec} in all seasons (Table 4) and SDSM next best. The results were not so unanimous for the RPS_{dry} and RPS_{wet}. The MOFRBC performed overall best concerning RPS_{dry}, and SDSM best for RPS_{wet}. The analogue methods outperformed the more complicated methods during winter in terms of the STARDEX variables (Table 5). The PCA provided the driest

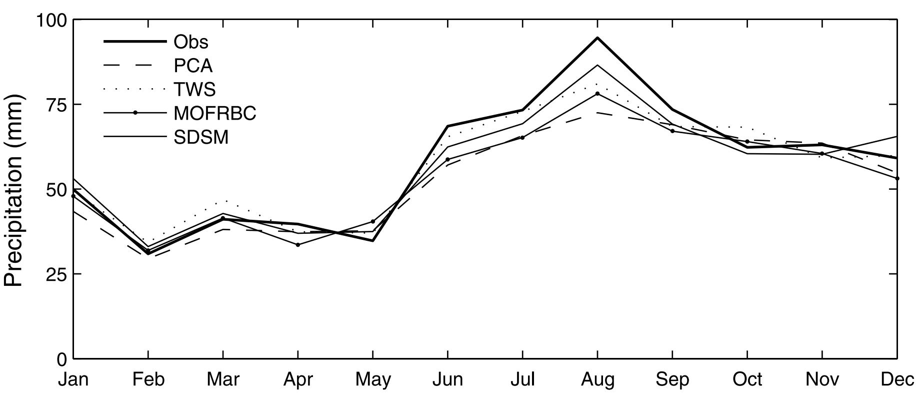

Fig. 7. Observed (Obs) and downscaled precipitation of monthly precipitation amounts averaged over seven precipitation stations in south-central Sweden for the validation period 1979–1994 with the four different downscaling method; the analogue methods (PCA and TWS), the conditional-probability method (MOFRBC), and the regression method with a weather generator (SDSM)

Table 4. Ranked-probability scores of downscaled precipitation during the validation period 1979–1993, averaged from seven precipitation stations in south-central Sweden, for four downscaling methods; the analogue methods (PCA and TWS), the conditional-probability method (MOFRBC), and the regression method (SDSM). Relative ranks are given as superscripts

	PCA	TWS	MOFBRC	SDSM
Winter (DJF)				
RPS_{prec}	7.8^{3}	8.4^{4}	5.3^{1}	6.4^{2}
RPS_{dry}	13^{1}	14^{3}	13^{1}	14^{3}
RPS_{wet}	12^{3}	9.3^{1}	13^{4}	11^{2}
Spring (MAM)				
RPS_{prec}	6.8^{3}	8.2^{4}	4.5^{1}	4.8^{2}
RPS_{dry}	13^{3}	14^{4}	11^{1}	11^{1}
RPS_{wet}	6.7^{3}	8.2^{4}	5.6^{3}	5.9^{2}
Summer (JJA)				
RPS_{prec}	9.2^{3}	12^{4}	6.1^{1}	6.4^{2}
RPS_{dry}	9.3^{2}	13^{4}	9.1^{1}	10^{3}
RPS_{wet}	14^{4}	12^{2}	13^{3}	9.5^{1}
Fall (SON)				
RPS_{prec}	9.8^{3}	11^{4}	6.4^{1}	6.9^{2}
RPS_{dry}	10^{4}	9.4^{3}	8.1^{2}	7.6^{1}
RPS_{wet}	11^{4}	9.7^{2}	10^{3}	8.7^{1}

climate and the best result for precipitation amount on a wet day and second best 5-day maximum in winter. It was notable that MOFRBC managed to downscale maximum length of dry spells best for all seasons except summer, otherwise the method performed poorly during winter and spring for the STARDEX indices. All methods, except MOFRBC, underestimated precipitation amount for a rainy summer day. This method was also the best overall method for summer.

The MOFRBC downscaled the STARDEX variables best for the wettest summers, except for maximum length of dry spell (Table 6). The TWS method performed best overall during dry summers. No method captured the magnitude of difference between wet and dry climate conditions very well, but a one-sided rank-sum test (Mann-Whitney) rejected the hypothesis that 5-day maximum rainfall came from the same population for SDSM and MOFRBC on the significance level 0.01.

6. Discussion

This study showed significant differences in how seasonality was processed by different statistical downscaling methods. Furthermore, downscaling ability was seasonally dependant. We suggest that important differences are related to (1) the geographical areas that determine precipitation in winter and summer, and (2) how this spatial information is used by the different methods.

6.1 Spatial correlations

The PCA method required large areas to optimise results for summer precipitation (Fig. 5). This was expected because the leading PCs had important modes around Iceland (Fig. 2). This, in turn, may be an effect resulting from the high variation in the MSLP field west of Sweden. Both summer and winter PC_1 were centred in the North Atlantic Ocean between Iceland and Norway and contributed to about 1/3 of the variance. The centre shifted somewhat towards the south in the winter both for PC_2 and PC_3. This indicated slightly different circulation patterns for winter and summer months. A good downscaling model should reflect if the daily precipitation is governed by regional pressure fields. The optimum PCA time window was almost 7 times larger in summer than in winter. This indicated that more data is needed during summer to correctly capture the variability.

The circulation patterns (CPs) in MOFRBC (Fig. 3) represent another decomposition of the MSLP-anomaly field (Fig. 2) optimised to give differences in precipitation. CP_6 and CP_{12} were the wettest patterns and, although they were not very frequent (about 5% each); they provided the patterns for which extreme precipitation amounts were expected. These patterns had similarities with PC_2 (Fig. 2). The most frequent pattern, CP_3, had anticyclonic centres close to the station area and a strong resemblance to the first principal component of the MSLP (Fig. 2). This is important if the classification done in the MOFRBC method is to possess physical credibility i.e. the composite anomaly patterns must correspond to distinguishable weather situations.

The correlation patterns between MSLP and precipitation in the SDSM model indicated that the most important non-delayed MSLP-grid points were located southwest of the station area. The summer precipitation is mainly driven by local convective circulation, but Fig. 6 also shows a westerly influence. The winter precipitation had

Table 5. Evaluation variables for observed (average and standard deviation) and downscaled precipitation during the validation period 1979–1993, averaged from seven precipitation stations in south-central Sweden, for four downscaling methods; the analogue methods (PCA and TWS), the conditional-probability method (MOFRBC), and the regression method (SDSM). The top number in each pair gives the variable value; the lower number (in italics) gives the residual-function value

	Obs ± std. dev.	PCA	TWS	MOFRBC	SDSM
Winter (DJF)					
Average-wet-day amount (mm/day)	3.40 ± 0.56	3.49^1	3.58^2	3.82^4	3.69^3
RF_{amount}		*<0.1*	*0.18*	*0.41*	*0.29*
90th-percentile-of-rain-day amounts (mm/day)	8.65 ± 1.9	9.17^3	8.99^2	9.63^4	8.82^1
$RF_{90\%}$		*0.52*	*0.34*	*0.99*	*0.17*
Greatest 5-day total rainfall (mm)	53	65^4	64^3	61^1	62^2
$RF_{\max}$		*11*	*11*	*8.0*	*8.3*
Maximum length of dry spell (days)	21	18^4	22^2	21^1	24^3
RF_{dry}		*3.6*	*0.9*	*3.8*	*2.4*
Spring (MAM)					
Average-wet-day amount (mm/day)	3.52 ± 0.73	3.48^1	3.45^2	3.85^4	3.59^3
RF_{amount}		*<0.1*	*<0.1*	*0.34*	*<0.1*
90th-percentile-of-rain-day amounts (mm/day)	8.86 ± 2.2	8.71^3	8.90^1	9.65^4	8.80^2
$RF_{90\%}$		*0.15*	*0.04*	*0.80*	*0.06*
Greatest 5-day total rainfall (mm)	51	58^3	53^1	63^4	57^2
$RF_{\max}$		*6.8*	*1.8*	*12*	*5.8*
Maximum length of dry spell (days)	24	25^2	28^3	24^1	28^4
RF_{dry}		*0.25*	*3.3*	*0.4*	*4.0*
Summer (JJA)					
Average-wet-day amount (mm/day)	5.78 ± 1.2	5.43^3	4.96^4	5.91^1	5.44^2
RF_{amount}		*0.35*	*0.82*	*0.12*	*0.33*
90th-percentile-of-rain-day amounts (mm/day)	14.5 ± 3.9	14.2^1	12.5^4	15.1^2	13.5^3
$RF_{90\%}$		*0.26*	*1.9*	*0.62*	*0.98*
Greatest 5-day total rainfall (mm)	109	99^4	100^3	104^2	109^1
$RF_{\max}$		*9.6*	*9.3*	*5.3*	*0.26*
Maximum length of dry spell (days)	16	18^1	21^3	19^2	26^4
RF_{dry}		*2.0*	*5.2*	*3.2*	*10*
Fall (SON)					
Average-wet-day amount (mm/day)	4.53 ± 1.15	4.45^2	4.48^1	5.00^4	4.44^3
RF_{amount}		*<0.1*	*<0.1*	*0.47*	*<0.1*
90th-percentile-of-rain-day amounts (mm/day)	11.4 ± 3.0	10.9^2	11.1^1	12.5^4	10.7^3
$RF_{90\%}$		*0.54*	*0.33*	*1.1*	*0.73*
Greatest 5-day total rainfall (mm)	79	97^4	94^3	86^2	82^1
$RF_{\max}$		*18*	*15*	*7.5*	*2.7*
Maximum length of dry spell (days)	17	15^4	18^2	17^1	19^3
RF_{dry}		*2.4*	*1.1*	*<0.1*	*1.6*

an area of influence southwest of the study area, indicating a strong relationship between westerly winds and precipitation during winter. It can be noted that this area was included in the optimum area for the winter TWS simulations.

6.2 Seasonal variations

All methods generated seasonal precipitation variations similar to the observed values (Fig. 7). Precipitation during winter, spring and autumn seasons was generally well downscaled by all methods, but summer precipitation was underestimated. This was especially the case for August, and SDSM performed best during this month. This suggested that SDSM best captures local convective precipitation, which is not indicated in the STARDEX variables, but can be seen in the precipitation probabilities. MOFRBC captured STARDEX variables best for summer, but underestimated precipitation totals. The TWS method had constantly higher precipitation totals

Table 6. Evaluation variables for observed (average and standard deviation) and downscaled precipitation for the five driest and the five wettest summers, averaged from seven precipitation stations in south-central Sweden, for the four downscaling methods

	Obs ± std. dev.	PCA	TWS	MOFRBC	SDSM
The 5 driest summers					
Average-wet-day amount (mm/day)	4.42 ± 1.2	5.00	4.88	5.72	5.08
RF_{amount}		*0.58*	*0.46*	*1.3*	*0.66*
90^{th}-percentile-of-rain-day amounts (mm/day)	11 ± 4.5	13	12	15	13
$RF_{90\%}$		*1.5*	*1.2*	*3.7*	*1.4*
Greatest 5-day total rainfall (mm)	57	73	64	63	64
$RF_{\max}$		*16*	*7.0*	*6.7*	*7.1*
Maximum length of dry spell (days)	19	14	20	17	21
RF_{dry}		*5.6*	*0.7*	*2.4*	*2.1*
The 5 wettest summers					
Average-wet-day amount (mm/day)	$6.45 \pm 1.1^{**}$	4.80	5.55^{*}	6.13^{*}	5.35
RF_{amount}		*1.6*	*0.89*	*0.32*	*1.1*
90^{th}-percentile-of-rain-day amounts (mm/day)	$16 \pm 3.9^{*}$	12	14	16	14
$RF_{90\%}$		*4.0*	*1.7*	*<0.1*	*2.2*
Greatest 5-day total rainfall (mm)	101^{**}	69	75^{*}	89^{**}	75^{**}
$RF_{\max}$		*33*	*27*	*13*	*26*
Maximum length of dry spell (days)	11^{*}	14	13	15	15
RF_{dry}		*2.8*	*2.0*	*4.2*	*3.8*

Significance levels of rejection *0.05 and **0.01 of the one-sided Mann-Whitney test with "H_0: The values are equal"

in summer compared to MOFRBC but much lower STARDEX variable values. This paradox was explained by the number of wet days. The probability of a wet day was 0.46 with the TWS method and 0.38 with the MOFRBC method, making the total summer rainfall greater with the TWS method. The STARDEX variables were designed to give emphasis on extreme events rather than monthly totals, therefore the inability of the MOFRBC method to capture monthly total may be a consequence of too few wet days. The MOFRBC was designed to capture extreme events in this study, and could perhaps downscale monthly totals better if the weather patterns were optimised towards this.

All methods except TWS overestimated precipitation amounts and maximum precipitation over 5 consecutive winter days. TWS performed best overall in winter. This result, along with the smaller window for the TWS method in winter, indicated that small areas are needed to capture the circulation that primarily governs daily winter precipitation over the study area.

The difference between wet and dry summers is apparent in the observed data (Table 6). However, no method could reproduce the large differences. A Wilcoxon-rank-sum test was rejected on the significance level 0.01 that the 5-day maximum for MOFRBC and SDSM came from the same population and on the 0.05 level for TWS. The test was also rejected for MOFRBC and TWS for wet-day amount on the 0.05 level. This may indicate that these methods are sensitive to different climate regimes, but this needs to be more rigorously tested. It is however promising that the signal is present. PCA was not sensitive to different climate regimes.

6.3 Downscaling skill

MOFRBC and SDSM performed well in comparison to the benchmark methods over the seasons (Table 4), especially when looking at RPS_{prec} where the analogue methods were clearly outperformed. The MOFRBC captured the durability and distribution of dry spells overall best (Tables 4, 5). This indicates that large-scale circulation of predictor data are important for a good downscaling of dry patterns. MOFRBC tended to overestimate precipitation amount on a wet day and 5-day maximum whereas SDSM were closer to the observed values and had the lowest RPS_{wet} for summer and autumn. This indicates that local variations in the MSLP field are important for the variations in precipitation.

The different methods performed differently well depending on season, so which method is preferable? No method performed best in all seasons. The choice of method is instead dependant on the purpose of the study. The recommendation of the authors is to use a number of methods in the downscaling study in order to achieve diversity within the results. There is also the question as to whether or not the methods can be used to downscale precipitation in simulation of a future perturbed climate. Since the two analogue sampling methods only reproduce historical data the future precipitation and circulation patterns have to be assumed to have the same variability as the present in order to simulate precipitation in a perturbed climate. This also applies to the two other methods since no method is valid outside its calibrated interval, but at least the MOFRBC and SDSM possess the potential to produce good results in a different climate. The analogue methods can always serve as benchmarks to more complicated methods in downscaling studies (Zorita and von Storch, 1999).

In this study only MSLP was used as the predictor. One could argue that other large-scale parameters, alone or in combinations, such as geopotential heights, vorticities and humidity, may increase the performance of the methods. Introducing more predictors, however, increases the degrees of freedom. This increases difficulties for results interpretation and method optimisation. The purpose of this study was to compare the methods as objectively as possible. This boundary condition led to the selection of MSLP as predictor.

7. Conclusions

The SDSM and MOFRBC performed on par or better than the analogue benchmark methods in terms of reproducing precipitation distributions for all seasons. MOFRBC captured the RPS_{prec} best, but had a tendency to produce too wet patterns. The methods in the study performed differently depending on season concerning indices such as precipitation amount on a rainy day and greatest 5-day total rainfall. The analogue methods were better during winter and autumn; SDSM and TWS during spring; and MOFRBC during summer. Larger predictor areas were necessary for summer precipitation than for other seasons. This indicates that large-scale circulations are important for the downscaled persistence of weather patterns, especially dry spells. In contrast, the SDSM method applied using a smaller predictor area captured precipitation intra-annual variance and RPS_{wet} well. This indicates that local MSLP activities are important for precipitation patterns. The TWS, SDSM and MOFRBC methods captured differences between extreme climate situations better than PCA.

Acknowledgements

This study was made possible by internal funding from Uppsala University and external funding from the Swedish Natural Science Research Council (contract G5301-972) and the Swedish Research Council (contract 621-2002-4352).

References

Bardossy A, Duckstein L, Bogardi I (1995) Fuzzy rule-based classification of atmospheric circulation patterns. Int J Climatol 15: 1087–1097

Bardossy A, Plate EJ (1992) Space-time model for daily rainfall using atmospheric circulation patterns. Water Resources Res 28: 1247–1259

Barrow EM, Semenov MA (1995) Climate change scenarios with high spatial and temporal resolution for agricultural applications. Forestry 68: 349–360

Baur F, Hess P, Nagel H (1944) Kalender der Grosswetterlagen Europas 1881–1939. Bad Homburg, 35 pp

Beckmann BR, Buishand TA (2002) Statistical downscaling relationships for precipitation in the Netherlands and North Germany. Int J Climatol 22: 15–32

Biau G, Zorita E, von Storch H, Wackernagel H (1999) Estimation of precipitation by kriging in the EOF Space of the sea level pressure field. J Clim 12: 1070–1085

Conway D, Jones PD (1998) The use of weather types and air flow indices for GCM downscaling. J Hydrol 212–213: 348–361

Cubasch U, von Storch H, Waszkewitz E, Zorita E (1996) Estimates of climate changes in southern Europe using different downscaling techniques. Climate Res 7: 129–149

Enke W, Schneider F, Deutschländer T (2005) A novel scheme to derive optimized circulation pattern classifications for downscaling and forecast purposes. Theor Appl Climatol 82: 51–63

Epstein ES (1969) A scoring system for probability forecasts of ranked categories. J Appl Meteor 8: 985–987

Eriksson B (1983) Data Rörande Sveriges Nederbördsklimat. Normalvärden för Perioden 1952–1980 (Data concerning the precipitation climate of Sweden. Mean values for the period 1951–1980; in Swedish with English abstract). Swedish Meteorological and Hydrological Institute Norrköping 1983: 28, 92 pp

Halldin S, Gryning S-E, Gottschalk L, Jochum A, Lundin L-C, Van de Griend AA (1999) Energy, water and carbon exchange in a boreal forest landscape – NOPEX experiences. Agric Forest Meteorol 98–99: 5–29
Hanssen-Bauer I, Førland EJ (2000) Temperature and precipitation variations in Norway and their links to atmospheric circulation. Int J Climatol 20: 1693–1708
Hay LE, McCabe GJ, Wolock DM, Ayers MA (1991) Simulation of precipitation by weather type analysis. Water Resources Res 27: 493–501
Huth R, Kysely J (2000) Constructing site-specific climate change scenarios on a monthly scale using statistical downscaling. Theor Appl Climatol 66: 13–27
Johansson B, Chen D (2003) The influence of wind and topography on precipitation distribution in Sweden: Statistical analysis and modelling. Int J Climatol 23: 1523–1535
Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen 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) NCEP/NCAR 40-year reanalysis project. Bull Amer Meteor Soc 77: 437–471
Kilsby CG, Cowpertwaith PSP, O'Connell PE, Jones PD (1998) Predicting rainfall statistics in England and Wales using atmospheric circulation variables. Int J Climatol 18: 523–539
Murphy AH (1971) A note on the ranked probability score. J Appl Meteor 10: 155–156
Obled C, Bontron G, Garcon R (2002) Quantitative precipitation forecasts: a statistical adaptation of model outputs through an analogues sorting approach. Atmos Res 63: 303–324
STARDEX (2001) Statistical and regional dynamical downscaling of extremes for European regions, description of work. http://www.cru.uea.ac.uk/cru/projects/stardex/description.pdf Accessed 2004-02-17
Stehlik J, Bardossy A (2002) Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation. J Hydrol 256: 120–141
Stehlik J, Bardossy A (2003) Statistical comparison of European circulation patterns and development of a continental scale classification. Theor Appl Climatol 76: 31–46
Teweles S Jr, Wobus HB (1954) Verification of prognostic charts. Bull Amer Meteor Soc 35: 455–463
Wilby RL, Dawson CW, Barrow EM (2002) SDSM – a decision support tool for the assessment of regional climate change impacts. Environ Model Software 17: 147–159
Wilby RL, Hay LE, Leavesley GH (1999) A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River Basin, Colorado. J Hydrol 225: 67–91
Wilby RL, Wigley TML (1997) Downscaling general circulation model output: a review of methods and limitations. Progress Phys Geogr 21: 530–548
Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, Wilks DS (1998) Statistical downscaling of general circulation model output: A comparison of methods. Water Resources Res 34: 2995–3008
Wetterhall F, Halldin S, Xu C-Y (2005) Statistical precipitation downscaling in central Sweden with the analogue method. J Hydrol 360: 174–190
Xu C-Y (1999) From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. Progress Phys Geogr 23: 229–249
Xu C-Y, Seibert J, Halldin S (1996) Regional water balance modelling in the NOPEX area: development and application of monthly water balance models. J Hydrol 180: 211–236
Zorita E, Hughes JP, Lettenmaier DP, von Storch H (1995) Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. J Clim 13: 223–234
Zorita E, von Storch H (1999) The analogue method as a simple statistical downscaling technique: Comparison with more complicated methods. J Clim 12: 2474–2489

Authors' address: Fredrik Wetterhall (e-mail: fredrik.wetterhall@hyd.uu.se), Sven Halldin, Chong-yu Xu, Air and Water Science, Department of Earth Sciences, Uppsala University, Villavägen 16, 752 36 Uppsala, Sweden.