

# Comparison and evaluation of statistical downscaling techniques for station-based precipitation in the Middle East

K. Lutz,<sup>a,\*</sup> J. Jacobeit,<sup>a</sup> A. Philipp,<sup>a</sup> S. Seubert,<sup>a</sup> H. Kunstmann<sup>a,b</sup> and P. Laux<sup>b</sup>

<sup>a</sup> Institute of Geography, University of Augsburg, Universitätsstr. 10, D-86135 Augsburg, Germany

<sup>b</sup> Institute for Meteorology and Climate Research IMK-IFU, Karlsruhe Institute of Technology, Kreuzackbahnstr. 19, D-82467 Garmisch-Partenkirchen, Germany

**ABSTRACT:** Several statistical downscaling techniques are intercompared and evaluated with respect to daily station-based precipitation in the eastern Mediterranean/Middle East region. The study introduces unconditioned and precipitation-conditioned SANDRA (Simulated ANnealing and Diversified RAndomization) cluster analysis (SCA) as new downscaling approaches and additionally uses the two widely used techniques of canonical correlation analysis (CCA) and multiple linear regression analysis (MR). For the precipitation-conditioned SANDRA cluster analysis different weights (percentages of contribution to the clustering) are evaluated. Furthermore, two different predictor combinations are used, a simple one only including mean sea level pressure (SLP), and a more complex one additionally including 500 hPa-geopotential heights, 500 hPa-vorticity and 1000 hPa-moisture flux. Analyses are carried out on a daily basis for the main rainy season from November to March for the period 1961–1990. It is shown that SLP, as single predictor, does not perform sufficiently well, but adding further predictors considerably improves model performance in terms of increased explained variance and model stability as well as reduced root mean square error (RMSE). From all selected techniques MR and CCA show the best performance for the SLP-based models, with comparable results for both techniques, whereas precipitation-conditioned SANDRA cluster analysis performs best when further predictors are included. Performance differences between all techniques are generally smaller than those for a particular technique using different predictor sets.

KEY WORDS statistical downscaling; station-based precipitation; Middle East

## 1. Introduction

Sufficient freshwater availability is a central prerequisite for agricultural and industrial development in the water-scarce environment of the eastern Mediterranean and Middle East (EM/ME) regions (Suppan *et al.*, 2008). Especially in the semiarid to arid ME region, water availability may be seriously affected by climate change-induced small changes of the prevailing seasonal precipitation patterns. Water scarcity has often caused conflicts between the adjacent countries of Israel, Palestine, Jordan, Syria, and Lebanon. The need for reliably predicting the future water availability is obvious. This requires a better understanding of recent precipitation characteristics and the development of methods to project future precipitation on regional scales.

However, the performance of general circulation models (GCMs) applied for impact studies at local scales is poor due to their coarse resolutions (e.g. Giorgi *et al.*, 2001). To obtain high-resolution climate change

scenarios, two main approaches have been developed: (1) *dynamical downscaling* based on regional climate models (e.g. Giorgi *et al.*, 2001), a very CPU expensive approach, and (2) *statistical downscaling* based on empirical relationships between appropriate large-scale meteorological variables (predictors) and the variable of interest (predictand) on a regional or local scale.

The focus of this study is on the comparison of the performance of different statistical downscaling techniques. Statistical relationships are established based on a training set (calibration period), verified during an independent verification period, and are applied for projections if necessary assumptions are fulfilled (IPCC, 2007). The various statistical downscaling methods can be classified into three main categories (Wilby *et al.*, 2004): (1) Weather Classification schemes (or Synoptic Downscaling), (2) Regression-based models, and (3) Weather Generators. The Weather Classifications can generally be grouped into subjective, semi-objective and objective approaches (Huth *et al.*, 2008; Jacobeit, 2010). Prominent examples for Weather Classification schemes are cluster analysis combined with the analogue method (Zorita and von Storch, 1999), Lamb Weather Type classification

\* Correspondence to: K. Lutz, Institute of Geography, University of Augsburg, Universitätsstr. 10, D-86135 Augsburg, Germany.  
E-mail: karin.lutz@geo.uni-augsburg.de

combined with Markov Chains (Goodess and Palutikof, 1998), objective fuzzy rule-based classification (Bárdossy *et al.*, 2002), simulated annealing classification (Philipp *et al.*, 2007) or hidden Markov models (Kioutsoukis *et al.*, 2008). Regression-based models include simple or multiple regressions (Jacobeit, 1996; Huth, 2004; Hertig and Jacobeit, 2008) and canonical correlation analysis (Barnett and Preisendorfer, 1987; Busuioc *et al.*, 2008; Hertig and Jacobeit, 2008). Weather Generators use stochastic models based on either Markov chains or series-based approaches (Huth *et al.*, 2001; Buishand *et al.*, 2004). As each of these statistical downscaling techniques has its advantages and disadvantages (Wilby *et al.*, 2004) and their performance depends on the season and region of interest, several, but still few comprehensive studies for comparing statistical downscaling methods have been performed. Wilby *et al.* (1998) used weather generators, resampling methods and artificial neural networks (ANN) for downscaling daily precipitation at six US regions. Zorita and Storch (1999) compared the analogue method to several other statistical downscaling techniques including linear models, classification methods and neural networks. Haylock *et al.* (2006) applied CCA, four techniques employing ANNs and a resampling method for downscaling seasonal indices of heavy precipitation for two station networks in the northeast and southeast of England. However, none of the studies clearly favours a specific methodology.

### 1.1. Downscaling studies for the Middle East region

Most of the existing research works on climate variability and change over Europe include only partially the Mediterranean basin and the ME region as the southernmost part of the research domain (Li *et al.*, 2006). Only few dynamical downscaling studies were conducted, which were centred on the Mediterranean basin (e.g. Giorgi *et al.*, 2004a, 2004b). Regional climate simulation results for the ME region are expected to be biased due to marginal effects. Suppan *et al.* (2008) and Kunstmann *et al.* (2007) performed one-way nested dynamical downscaling in 18 km spatial resolution to estimate the expected climate change impact on water availability in the ME using the regional climate model MM5 and the A2 and B2 scenario-based ECHAM4 data.

Hertig and Jacobeit (2008) used two statistical downscaling methods, namely canonical correlation analysis (CCA) and multiple regression analysis (MR) to assess the expected Mediterranean precipitation changes for the period 1990–2100 under increased greenhouse gas conditions. For the ME region, they found mainly negative precipitation changes for the rainy season ranging from October to May.

Several weather classifications were developed and applied for the ME. The first classification for the ME region was performed by Koplowitz (1973) using synoptic pressure field patterns. Ronberg (1984) used radiosonde and surface data for his classification over the southeastern ME coastal plain (Israel) for the October–April period and identified 18 weather patterns,

which were grouped into 4 types: *Red Sea Trough types*, *Transitional types*, *Sharav-like types*, and *Stormy types*. Most of the classification results were linked to hydrometeorological features, such as wind energy regimes (e.g. Alpert *et al.*, 1987; Shafir *et al.*, 1994), floods (Kahana *et al.*, 2002) and rainfall (e.g. Kutiel and Paz, 1998; Ribera *et al.*, 2000; Maheras *et al.*, 2001; Zangvil *et al.*, 2003; Alpert *et al.*, 2004; Ziv *et al.*, 2006). Zangvil *et al.* (2003) identified typical SLP distributions for 4 GHP500 prototype patterns, occurring predominantly on major rainy days. Ziv *et al.* (2006) found an upper level trough extending from eastern Europe towards the ME region which is linked to December–February rainfall in the northern part of Israel. Alpert *et al.* (2004) identified the Red Sea Trough to be responsible for a decreasing rainfall trend in the ME region. In an earlier study, Alpert *et al.* (1990) referred rainfall occurrence in the ME mainly to the passages of extratropical cyclones, the Cyprus lows. On the basis of the semi-objective classification of Alpert *et al.* (2004), Saaroni *et al.* (2010) analysed the inter-annual variation and spatial distribution of rainfall with respect to variations in the occurrence of typical synoptic systems. The occurrence of rainfall events during the months November–March was found to be highly correlated with the occurrence of Cyprus lows. They also point out that intensity and spatial distribution of rainfall is highly sensitive to location and gradient of the Cyprus lows.

Comparisons of different circulation pattern classifications have been conducted by several authors, however, mostly constrained for western-central Europe (e.g. Jones *et al.*, 1993; Buishand and Brandsma, 1997; Stehlík and Bárdossy, 2003; Anagnostopoulou *et al.*, 2004).

Laux *et al.* (2008a) compared the performance of two objective and one semi-objective weather pattern classifications for precipitation simulation in the ME region dependent on the season. They concluded that the applied semi-objective classification, partly based on expert knowledge performs slightly better than the objective ones with no *a priori* knowledge required. Dependencies in the occurrence of distinct weather patterns between the different classifications were identified. In another study, Laux *et al.* (2008b) performed frequency analysis of weather patterns causing extreme wet and droughty conditions for the past (1961–1990) and future time slice (2011–2040) based on the A1B scenario and ECHAM5 data. They concluded that no remarkable differences in the frequencies of extremes are expected for the period 2011–2040.

### 1.2. Aims of this study

The aims of this study are:

- i) To compare the performance of three different statistical downscaling techniques in terms of rainfall simulation for the ME region. The techniques include two widely used regression-based, MR and CCA (Section

3.1), as well as two variants of a novel synoptic downscaling technique, unconditioned and precipitation-conditioned Simulated ANnealing and Diversified RAndomization clustering (SANDRA) cluster analysis (SCA) (Section 3.2). Precipitation-conditioned cluster analysis includes the predictand precipitation in the clustering whereas unconditioned cluster analysis only includes the predictor variables.

- ii) To evaluate the effect of the selected predictor variables on the modelling results. In a first step, only sea level pressure (SLP) is used as predictor for models whose performance is evaluated in Section 4.2. In a second step, further atmospheric predictors, such as geopotential heights, relative vorticity and moisture flux of different levels, are included in the models and the improvement of the prediction skills is analysed (Section 4.3).

## 2. Data

Different datasets extracted from the National Center for Environmental Prediction (NCEP) - National Center for Atmospheric Research (NCAR) reanalysis project (Kalnay *et al.*, 1996; Kistler *et al.*, 2001) are used in this study. These include SLP, geopotential heights ( $HGT$ ) of the 500 hPa level as well as specific humidity ( $SF$ ) and the  $u$ -wind component ( $u$ ) of the 1000 hPa level. Vorticity ( $VORT$ ) and moisture flux ( $MF_u$ ) are calculated as follows:

$$VORT = (HGT_{i-1,j} + HGT_{i+1,j} + HGT_{i,j-1} + HGT_{i,j+1}) - 4HGT_{i,j} \quad (1)$$

$$MF_u = SF \cdot u \quad (2)$$

where indices  $i$  and  $j$  indicate east-west and north-south labelling of gridpoints. The six-hourly data are available back to 1948 on a global  $2.5^\circ \times 2.5^\circ$  grid. The period of 1961–1990 is selected and daily means are calculated. The selected grid spans between  $22.5^\circ\text{N}$  and  $42.5^\circ\text{N}$ , and between  $15^\circ\text{E}$  and  $50^\circ\text{E}$ , comprising 135 grid points (Figure 1).

Sea level pressure,  $HGT$  and  $VORT$  of the 500 hPa level and moisture flux of the 1000 hPa level are chosen as large-scale atmospheric predictors for the downscaling

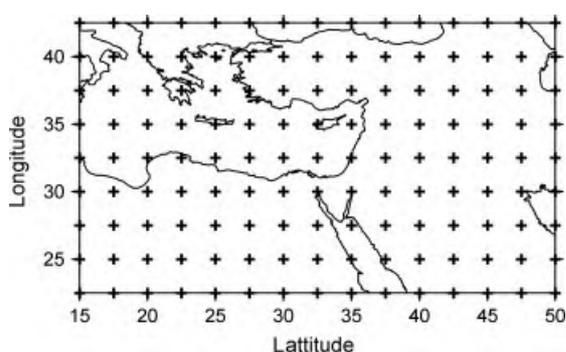


Figure 1. Location and extension of the predictor domain.



Figure 2. Location of stations used in this study. For details see Table I. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

models. Selection is based on previous experiments with various variables and different atmospheric levels.

Daily precipitation station data with different time series length is supplied by the GLOWA Jordan River project (<http://www.glowa-jordan-river.de>). These include various meteorological observation stations in Israel and Jordan. Owing to large gaps in most of the precipitation time series, only 26 stations in Israel and western Jordan that cover the period of 1961–1990 could be used as predictands (Figure 2 and Table I).

The study focuses on the main rainy season between November and March, when the monthly rainfall exceeds 10% of the long-term annual precipitation sum (given in Table I). All analyses for the study period 1961–1990 are carried out on a daily basis, separately for each month, as pre-analyses have shown considerable month-to-month variations in precipitation statistics that can only be captured in separate analyses.

## 3. Statistical downscaling techniques

### 3.1. Regression models

As a first step, orthogonally (Varimax) rotated principal component analysis (PCA, e.g. Preisendorfer, 1988; von Storch and Zwiers, 1999) was applied to station-based precipitation and to all predictor fields in order to reduce dimensions and to remove linear dependencies between

Table I. List of stations used in this study with latitudes and longitudes (in degrees), PC number (for PCA of station data see Section 3.1) and mean annual precipitation (mm/a). Representative stations (see section 3.1) are greyed out.

Station	Name	Lat.	Lon.	PC	mm/a
1	Amman airport	31.97	35.60	1	264
2	Deirabisaid	32.50	35.68	1	423
3	Deirallaagr station	32.20	35.60	1	277
4	Ennueiyime	32.42	35.92	1	313
5	Hartha	32.70	35.85	1	427
6	Irbidagr station	32.53	35.85	1	467
7	Naur	31.87	35.83	1	472
8	Ramtha boys school	32.57	36.02	1	264
9	Rihab	32.32	36.10	1	214
10	Qiryant-Anav	31.80	35.12	1	704
11	Jerusalem	31.76	35.21	1	550
12	Ramat-David	32.67	35.18	2	539
13	Kebutzat	32.75	35.55	2	419
14	Kefar Blum	33.18	35.60	2	532
15	Kefar Gilad	33.23	35.56	2	811
16	Yiron	33.06	35.45	2	794
17	Eilon	33.05	35.21	2	795
18	Har Kenaan	32.96	35.50	2	699
19	Ainbisas	31.20	35.67	3	305
20	Hemud	31.30	35.80	3	291
21	Ta le	30.83	35.60	3	267
22	Beer-Sheva	31.23	34.78	3	207
23	Sede	30.86	34.78	3	114
24	Dorot	31.33	34.66	3	356
25	Qiryant-Shaut	32.11	34.82	4	569
26	Tel Aviv	32.09	34.77	4	518

variables. S-mode PCAs were carried out separately for each of the selected months (Nov–Mar) and the five different calibration periods (Section 3.3) within the study period 1961–1990, to receive spatial centres of variation and associated time coefficients (PCs) for each input predictor field, each month and each calibration period. Independent PC time series for the corresponding verification periods (Section 3.3) are calculated using calibration PC loadings (defining spatial patterns based on weighted eigenvector components) being independent from verification data. For the station data, S-mode PCA of daily precipitation anomalies is used to divide the 26 stations into 4 groups. According to the highest PC loadings, one station of each group is then selected as representative of its group. These are Jerusalem, Kefar Gilad, Beer-Sheva, and Tel Aviv (greyed out in Table I).

All PCAs are based on the correlation matrix of the input variables. The so-called dominance criterion (Jacobbeit, 1993) determines the number of PCs to be extracted. Thus, for each extracted PC there has to exist one dominant loading being greater than all other PC's loadings, reflecting both the PC's relevance and its prominence compared to other PCs. To strengthen this constraint, for this study, the difference between the highest and the second highest value has to be three standard deviations of the loading pattern.

Depending on the analysed months, SLP fields are reduced to 5–6 PCs with cumulative explained variances

(EVs) of 89–92%, *HGT* data to 5 PCs (EVs of 91–92%), *VORT* to 13–19 PCs (EVs of 84–92%) and *MF<sub>u</sub>* to 10–14 PCs (EVs of 84–89%).

Subsequently, two classical downscaling techniques based on transfer functions, i.e. (1) CCA, (e.g. Barnett and Preisendorfer, 1987), and (2) MR, (e.g. Easterling, 1999; for a detailed description of both techniques, read von Storch and Zwiers, 1999, and Wilks, 2005) are used to link precipitation time series to the large-scale atmospheric circulation represented by their corresponding PC time series and establish predictor–predictand relationships for different calibration periods (Section 3.3).

CCA is carried out for each month and calibration period, processing all station time series in one analysis (Section 3.3). When using MR, a stepwise regression procedure (von Storch and Zwiers, 1999) that combines forward selection with backward elimination is applied to each rainfall station. To verify a significant influence of a predictor variable on the predictand, an *F*-test is applied on its corresponding regression coefficient, with significance levels of 0.05 for the intake and 0.10 for an elimination of a variable from the regression equation.

The results of CCA and MR are then used to assess the response of station-based precipitation to changes in large-scale circulation. Precipitation time series are modelled using output PCs and CCAs or regression coefficients, respectively. In addition, time series of each station modelled from CCA output need to be re-scaled and centred using mean and standard deviation from the corresponding calibration time series. For both methods, two kinds of predictor sets were used: one set consisting of SLP as a single predictor variable, the other set consisting of multiple variables, i.e. SLP, *HGT*, relative *VORT* and *MF<sub>u</sub>* of selected levels.

Instead of calibrating and verifying just one model for each application, this is done – slightly modified compared to Hertig and Jacobbeit (2008) – for an ensemble of several models with different calibration periods (Section 3.3) in order to verify the statistical relationships and to consider possible non-stationarities. Cross-validation allows to produce more reliable results and to apply those to an independent dataset.

### 3.2. Synoptic downscaling

Synoptic downscaling is based on the assumption that a certain circulation or weather pattern accounts for a particular weather condition. For the classification of large-scale circulation, a variety of methods exist. Evaluation and comparison of different circulation type classifications is, for example, performed within the COST Action 733 ‘Harmonisation and Applications of Weather Types Classifications for European Regions’ (Beck and Philipp, 2010). The simplest way to use circulation types for downscaling is the analogue method (Zorita and Storch, 1999). For each type, an analogue is looked up in historical observation data which is connected to certain weather conditions. These are then transferred to the circulation type.

In this study, a special type of cluster analysis developed by Philipp *et al.* (2007) called SANDRA is applied. For a detailed description, refer Philipp *et al.* (2007, 2010). The selection of this classification method is based on the improved performance of the simulated annealing algorithm compared to conventional *k*-means. The latter has no strategy to avoid local optima in the optimisation function (Philipp *et al.*, 2007, 2010; Fereday *et al.*, 2008) whereas SANDRA is able to approximate the global optimum.

SANDRA allows for the contribution of several large-scale atmospheric variables to the classification. For that purpose, all variables need to be normalized separately before clustering.

As circulation pattern classification does not include any information about the local weather state (e.g. precipitation), it might be meaningful to incorporate this information in the classification procedure, i.e. to condition the classification by precipitation or other predictand variables. For this purpose, normalized precipitation time series are added to the normalized predictor time series and processed together. To account for different numbers of parameters (grid points vs stations in this case) and to control the influence of the predictor and predictand variables on the clustering, a so-called *weighted cluster analysis* is performed. The contribution of each variable to the clustering is expressed by a percental factor, the weight  $\lambda$ , which is applied to the squared Euclidean distance  $D$ :

$$D' (X'_i, \overline{X'_j}) = D (X_i, \overline{X_j}) \cdot \lambda / m \quad (3)$$

where  $X$  is any predictor or predictand variable contributing to the analysis,  $X'$  the corresponding weighted variable.  $\overline{X}$  ( $\overline{X'}$ ) is the unweighted (weighted) cluster centroid,  $i$  ( $j$ ) is the object (cluster centroid) number of  $X$  ( $\overline{X}$ ) and  $m$  is the number of parameters which is used for scaling.

$D$  ( $D'$ ) are the unweighted (weighted) Euclidean distances between  $X$  ( $X'$ ) and its corresponding cluster centroid  $\overline{X}$  ( $\overline{X'}$ ), which are used to assign objects (daily predictor pattern or predictand values) to their nearest cluster centroid. To calculate the overall Euclidean Distance, predictors' and predictands' individual Euclidean Distances are summed up. As can be seen from Equation (3), the weighted distance represents the unweighted distance first scaled by  $1/m$ , which yields an equal contribution of all variables to the clustering. The scaled distance is then multiplied by the user-defined weight  $\lambda$ .

As multiplying  $D$  by the factor  $\lambda/m$  is computationally expensive (the factor is applied every time the Euclidean distance is recalculated during clustering iterations), the observational data needs to be scaled and weighted, separately for any predictor and predictand prior to the actual clustering. The factors to be applied to the different datasets can be directly derived from the equation of the overall Euclidean distance  $D^w$  between object  $i$  and

its cluster centroid  $j$  which is the sum of  $n$  different predictors' and predictands' Euclidean distances:

$$\begin{aligned} D^w(i, j) &= \sum_{v=1}^n \sum_{p=1}^{m(v)} \left( X'_{ip(v)} - \overline{X'_{jp(v)}} \right)^2 = \\ &= \sum_{v=1}^n \sum_{p=1}^{m(v)} \left( X_{ip(v)} - \overline{X_{jp(v)}} \right)^2 \cdot \lambda_{(v)} \cdot 1/m_{(v)} = \\ &= \sum_{v=1}^n \sum_{p=1}^{m(v)} \left( X_{ip(v)} \cdot \sqrt{\lambda_{(v)} \cdot 1/m_{(v)}} - \overline{X_{jp(v)}} \right. \\ &\quad \left. \cdot \sqrt{\lambda_{(v)} \cdot 1/m_{(v)}} \right)^2 \end{aligned} \quad (4)$$

Solving Equation (4) leads to the direct weighting of the data:

$$X'_{(v)} = X_{(v)} \cdot \sqrt{\lambda_{(v)} \cdot 1/m_{(v)}} \quad (5)$$

where  $X'(v)$  ( $v = 1 \dots n$ ) is any predictor and predictand scaled by its corresponding number of parameters  $m_{(v)}$  and weighted by its corresponding individually defined  $\lambda_{(v)}$ , prior to clustering. Note that the square root needs to be applied to any scaling and weighting factor as it is squared later in the equation of the Euclidean distance.

Thus, a weighting value of one (or 100%) means that all variables equally contribute to the classification regardless of the number ( $m$ ) of parameters. For this study,  $\lambda$  of the predictor variables is set to 1 (100%) whereas a range of different values is evaluated for the precipitation weight (Section 4.1). Moreover, weighting is applied separately for the four representative stations to account for the varying cause of rainfall events by different weather patterns in different regions.

Furthermore, the number of clusters has to be predefined. The estimation of an appropriate number is done by *t*-mode PCA assuming that the number of dominant patterns constitutes the number of patterns that are needed to represent the majority of variance within the set of predictor fields. Therefore, for each month in the study period 1961–1990, separate analyses were carried out for SLP (yielding a number of 7–9 clusters) as well as for combined analyses with all predictor variables (16–22 clusters).

For this study, four types of classifications are carried out for each month and different calibration periods (Section 3.3): a SLP-based approach, a multiple-variable-classification using SLP, *HGT*, *VORT* and *MF<sub>u</sub>*, and also weighted variants of both. Result of each classification is a time series that gives the cluster number for every day. A predictor-predictand relationship is then established by averaging precipitation of all the corresponding days of a cluster or circulation type, receiving a mean precipitation value for every type. To validate the model, calibration and verification predictor data is assigned to the most similar circulation type, which is represented by its centroid (i.e. mean cluster pattern), using the squared Euclidean distance. Subsequently, a precipitation time series is generated by inserting the circulation type specific mean

precipitation value. In the case of weighted cluster analysis, predictand data (precipitation) is excluded during validation and only predictor data (circulation variables) is assigned to the classification centroids. Therefore, with increasing weight, Euclidean distances deviate more and more from the distances calculated only on predictor data (as in case of validation) and an increasing number of days is assigned to a different cluster compared to the clustering procedure. These different assignments have to be considered when establishing the downscaling models. It is recommended to choose the lower weight when model skills are similar as less different assignments occur and models are more stable.

As for the transfer function models (Section 3.1), not only is just one model calibrated and verified for each type of classification and month, but an ensemble of several models with different calibration periods. This ensemble approach is described in the following section.

### 3.3. Calibration and verification – an ensemble approach

Within the 30-year study period from 1961 to 1990, a 10-year moving window selects 5 segments. The remaining 20 years are used to calibrate the model, which is verified within the corresponding 10 independent years (Table II). Thus, for each month, an ensemble of five different models is generated. For verification, observation-based and statistically derived precipitation are aggregated to monthly values and correlated with each other, separately for each ensemble member. Thereafter, an ensemble mean value and standard deviation of the Pearson correlation coefficients of all members is calculated (Figures 6 and 8). The standard deviation characterizes the variation within the model ensemble and is referred to as Within-Ensemble Variance (WEV). Significance of the correlation coefficients is verified by Student's *t*-test. Since significance depends on the number of cases (here the number of monthly precipitation amounts), thresholds for calibration and verification periods differ. For the calibration period, coefficients greater than 0.4 are significant at the 90% level; for the verification period, the threshold is 0.55. To characterize each method's general performance by one figure, correlation coefficients were averaged over all four representative stations and five months, yielding a calibration and a verification correlation coefficient (Section 4.4). As a second skill score, Root Mean Square Error (RMSE) was used to detect considerable positive/negative deviations of the modelled

Table II. Calibration and verification periods for the five different model ensemble members m1 to m5.

Model	Calibration	Verification
m1	1971–1990	1961–1970
m2	1961–1965, 1976–1990	1966–1975
m3	1961–1970, 1981–1990	1971–1980
m4	1961–1975, 1986–1990	1976–1985
m5	1961–1980	1981–1990

time series from the observed ones in case of significant correlations. As for the correlation coefficients, overall RMSE values averaged over all four stations and five months were calculated for calibration and verification periods.

Cross-validation is often used to estimate the potential forecast skill (Michaelsen 1987; Huth 2002) considering bias resulting from model fitting and predictor screening, or for short study periods to account for non-stationarities in the statistical relationships between predictor and predictands (von Storch and Zwiers, 1999; Murphy 2000). As pre-analyses have shown considerable non-stationarities in the statistical relationships between circulation and precipitation, a cross-validation approach was also applied in the present study. The partitioning of the study period in different samples (shown above) allows the verification of the model's quality and transferability, leading to an unbiased estimate of potential predictability. In case of an unstable relationship, the model performance is more dependent on the selected period, whereas stable connections yield more consistent results across the whole ensemble of models. This is particularly important for the transferability of the statistical relationships to an independent dataset.

## 4. Results

### 4.1. Weighting experiments

As discussed in Section 3.2, simple circulation classification does not include any information about precipitation. To improve classification and to condition it by precipitation, rainfall station data was added and weighted by a percental factor to regulate its influence on the clustering. This is directly connected to the identification of an appropriate weight. Too small weights yield results that marginally differ from a pure circulation classification, and clusters correspondingly do not represent wet and dry patterns. On the other hand, too large weights will lead to a classification mainly driven by precipitation, and clusters correspondingly include highly

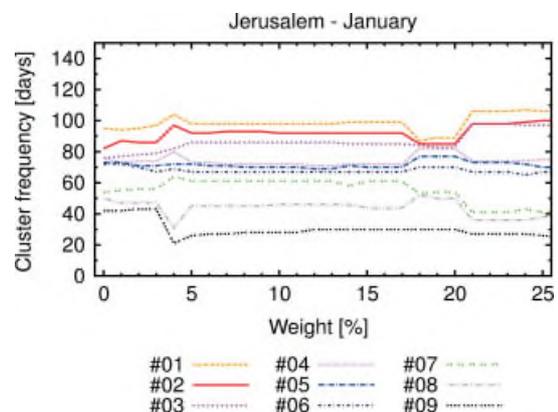


Figure 3. Cluster frequencies (in days) for different precipitation weights (x-axis) and a fixed cluster number of 9 for Jerusalem in January, ensemble member m3 (Section 3.3, Table II). This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

dissimilar circulation patterns. The latter also leads to a high number of wrong assignments during verification when only circulation data is used. To deal with this problem, a range of values between 1 and 25% is

applied and the results of the different weighted models are compared and evaluated. Figure 3 shows cluster frequencies, and Figure 4 centroids of the unweighted (left) and weighted models (right,  $\lambda = 5\%$ ) and

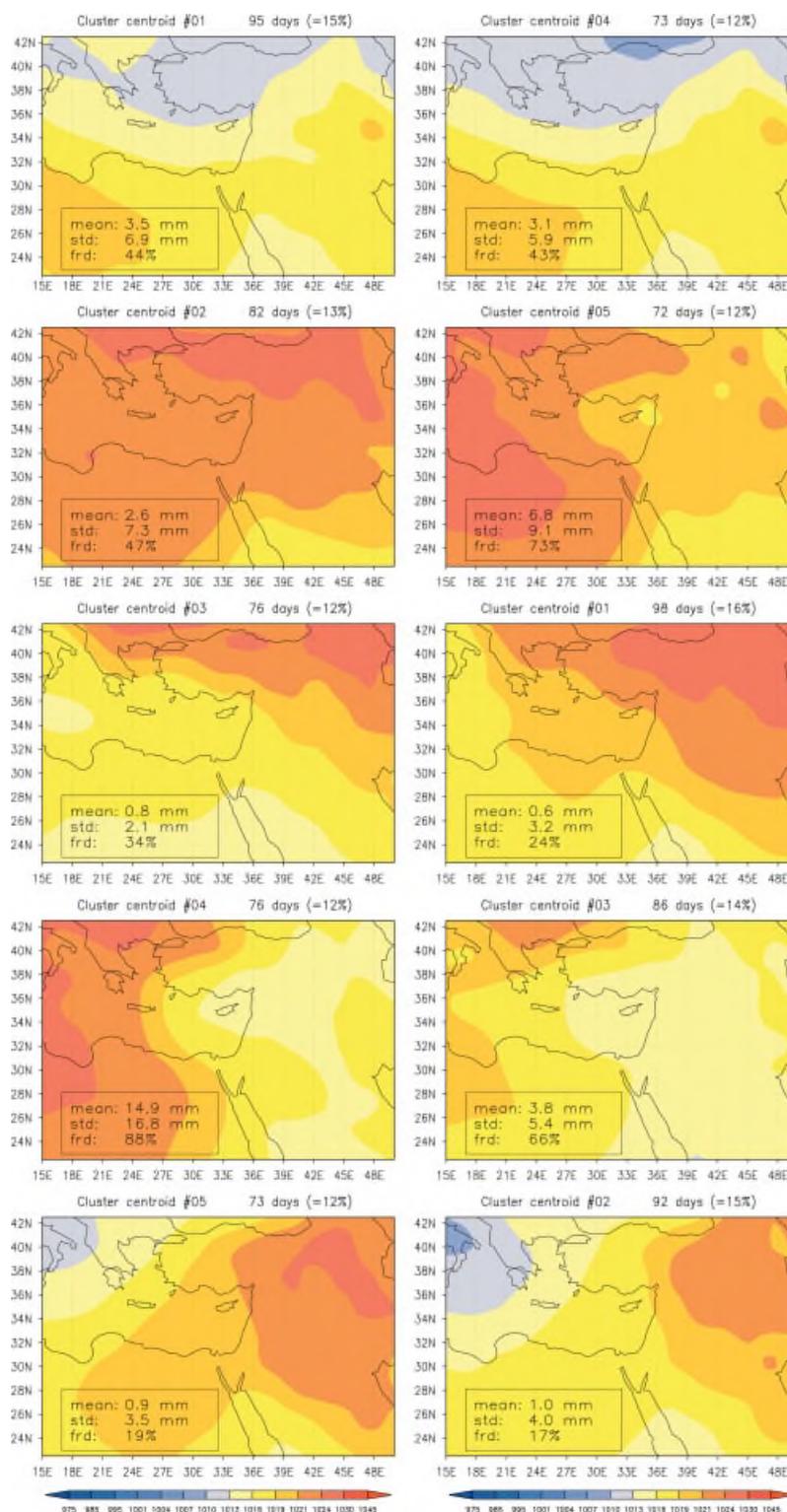


Figure 4. SLP Cluster centroids, corresponding pattern mean precipitation (mean), standard deviation (std) and frequencies of rain days (frd) (box, lower left corner) of an unweighted (left) and a weighted SANDRA cluster analysis (right,  $\lambda = 5\%$ ) conditioned on Jerusalem precipitation in January, ensemble member m3 (Section 3.3, Table II). Number of days assigned to each cluster is given on the top of each map. Spatial correlation is used to identify pairs so that each of the centroids from weighted SCA is sorted according to its most similar unweighted correspondent. Note that the cluster numbers remain unchanged and are set according to the centroids' frequency. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

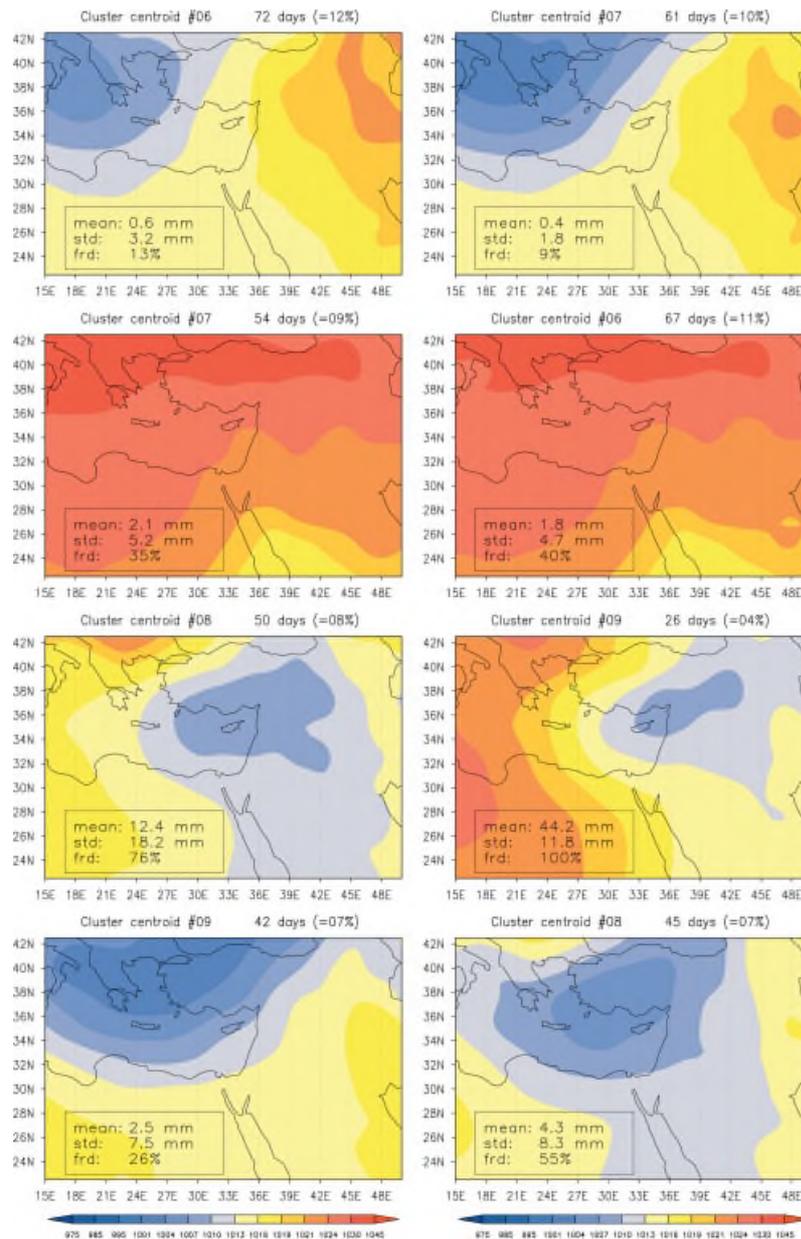


Figure 4. (Continued).

their corresponding mean precipitation values, standard deviations (which also serve as error estimate) and frequencies of rain days (box, left corner) for the example of a SLP-classification conditioned to Jerusalem precipitation in January for model period 3 (Section 3.3). In Figure 4, centroids from unweighted SANDRA Cluster Analysis (SCA) are sorted by frequency. Spatial correlation is used to identify pairs so that each of the centroids from weighted SCA is sorted according to its most similar unweighted correspondent. Note that the cluster numbers remain unchanged and are set according to the centroids' frequency. The unweighted classification (weight 0 in Figure 3) yields circulation types that differ marginally regarding their frequencies. Precipitation means are generally low, 3 clusters are associated with values below 1.0 mm. With increasing weight cluster frequencies first differ marginally, but at a weight of 4–5% a partitioning

into some clusters with frequencies below 50 days, some clusters with frequencies between 50 and 80 days, some others with frequencies above 90 days, and 1 cluster comprising only 26 days takes place. The most infrequent cluster also represents the circulation type that receives the highest amount of mean precipitation (44.2 mm). When further increasing the weight, only small changes occur that appear to be rather random than driven by precipitation. At a weight of 17% another partitioning takes place, thus it can be concluded that changes in frequencies with increasing precipitation weight occur stepwise. This was also found for other stations and months.

The highest mean precipitation amount (14.9 mm) for unweighted SCA is associated with lower pressure over Cyprus, Persia, Near East, and Arabia (cluster centroid #4) compared to the western oceanic regions. The SLP pattern yielding the second highest rainfall

amount (12.4 mm), centroid #8, shows low pressure above Cyprus and parts of continental near east, as well as above the Red Sea. In contrast to centroid #4, this pattern has a clear low-pressure centre over Cyprus, whereas in pattern #4, the low itself is shallow but the gradient to the western high is strong. Regardless of their depth, Cyprus lows are known to cause winter rainfall in Israel (Alpert *et al.*, 2004; Sarooni *et al.*, 2010, Section 4.4) with rainfall amounts depending on pressure gradients. As mean precipitation amounts are moderate and standard deviations are comparably high, it can be assumed that other pressure patterns than Cyprus lows which might be responsible for higher rainfall amounts, contribute to those two patterns which are associated with very small amounts of rain or no rain at all. Corresponding patterns may be Persian lows in the case of pattern #4 and Red Sea troughs in the case of pattern #8. For weighted SCA, similar patterns exist (Cluster centroids #3, #5, and #9; note that cluster number labels reflect frequency order). For unweighted cluster pattern #4, two similar patterns, #3 and #5, can be found. Their pressure gradients are less pronounced, rainfall means are lower (3.8 and 6.8 mm). Low pressure of pattern #9 (which corresponds to unweighted pattern #8) is more concentrated in northern parts and cannot be found above the Red Sea. The pressure gradient between high and low pressure is stronger in the weighted pattern #9. This pattern is also associated with the highest mean precipitation (44.2 mm), 100% rain days and a lower standard deviation (11.8 mm) than for the unweighted SCA pattern #8 (18.2 mm). This is also the maximum standard deviation value within the weighted classification being lower than the maximum in unweighted SCA. Therefore, weighted pattern #9 is also more homogeneous concerning precipitation.

For all different weights, downscaling models are established and evaluated for changes in performance with increasing weight. Therefore, mean ensemble correlation coefficients, difference between calibration and verification coefficients, and standard deviations are examined. Figure 5 shows an example of a SLP-classification conditioned by Jerusalem precipitation for January. For low weights, the calibration correlation coefficient first decreases and then increases to a peak at the weight of 4–5% where the continuously increasing verification correlation also shows a maximum with both values higher than the corresponding coefficients for unweighted clustering. For higher weights, there is no significant change in performance. The variance between the ensemble members shows higher fluctuations for low weights and lower ones for high weights. Choosing an appropriate weight is a compromise between the following criteria:

- Highest mean correlation coefficients
- Lowest difference between correlation coefficients for calibration and verification periods
- Lowest variability within the ensemble

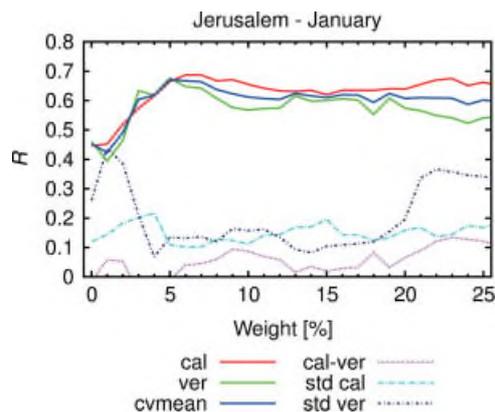


Figure 5. Ensemble mean correlation coefficients ( $R$ ) between observed and modelled January precipitation for different precipitation weights ( $x$ -axis) and a fixed cluster number of 9 for Jerusalem: calibration (cal), verification (ver), calibration-verification-mean (cvmean), calibration-verification-difference (cal-ver), calibration standard deviation (std cal), verification standard deviation (std ver). This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

Table III. Weights applied for precipitation-conditioned (weighted) SANDRA cluster analysis using sea level pressure as single predictor.

Station	Nov	Dec	Jan	Feb	Mar
Jerusalem	14	12	5	10	4
Kefar Gilad	13	0	25	12	9
Beer-Sheva	2	4	4	0	5
Tel Aviv	5	3	4	0	9

- Lowest weight in case of similarity of the first three criteria for different weights (Section 3.2).

A semi-objective procedure selects the most appropriate weights. For January and station Jerusalem, a weight of 5% is suggested. Examining all months and all four stations (Table III), the optimal weight is mostly below 20%, in many cases even below 10%, indicating a much higher variance within precipitation compared to SLP. Note that unweighted results (weight 0 in Table III) are kept and used for further considerations when weighting does not increase performance.

Weighting evaluation is also performed for the multiple-predictor-models. Weights are generally higher than those of the SLP-models because, in this case, precipitation is only 1 of 5 variables implying that a weight of 100% means an overall contribution of 20%. In 50% of the cases model performance is not improved by weighting; for Tel Aviv all months are best modelled by unweighted SCA (Table IV).

Using the same weight for all stations and months cannot be recommended as the classification is highly sensitive to differences in precipitation amounts and variability. To get the best result, an individual weight has to be chosen for each station and each month. This weight is applicable to any of the five sub-periods of the corresponding ensemble.

Table IV. Weights applied for precipitation-conditioned (weighted) SANDRA cluster analysis using sea level pressure, 500 hPa-geopotential heights, 500 hPa-vorticity, and 1000 hPa-moisture flux as predictors.

Station	Nov	Dec	Jan	Feb	Mar
Jerusalem	0	12	3	0	4
Kefar Gilad	0	11	0	16	10
Beer-Sheva	10	3	3	0	9
Tel Aviv	4	11	0	1	6

4.2. Models with sea level pressure as the only predictor

Differences in model performance arising from the application of different downscaling techniques using SLP as single predictor are displayed in Figure 6. Mean correlation coefficients ( $R$ ) vary strongly across techniques, months, and stations. For Jerusalem, coefficients are lowest for the drier transition months of November and March, the standard deviation indicates the existence of even negative values for the latter. Highest significant correlations can be found for the winter month December

(for thresholds refer Section 3.3). Within-ensemble variance (errorbars in Figure 6) is highest for March for all techniques and for the verification ensemble of February for the regression-based techniques, indicating non-stationarities in the predictor-predictand relationships. For Beer-Sheva, unweighted and weighted SCA fails to establish stable models in November, January, and March, showing mostly insignificant correlations and high WEV. In December and February, SCA performs better, with the weighted result being the best of all techniques. For Kefar Gilad and Tel Aviv all techniques (except for the unweighted SCA) perform reasonably well for the months of November–February with correlation coefficients above 0.5, while WEV is not satisfying. Considering RMSE (Figure 7), deviations between observed and modelled time series are highest in the wettest month of January and lowest in the drier transition month of March with lowest values for the driest station of Beer-Sheva. Different stations’ and months’ RMSE values, as well as different time periods, cannot be inter-compared as RMSE increases with increasing precipitation amounts. RMSE is also highly sensitive to errors in peak values as differences between observed and

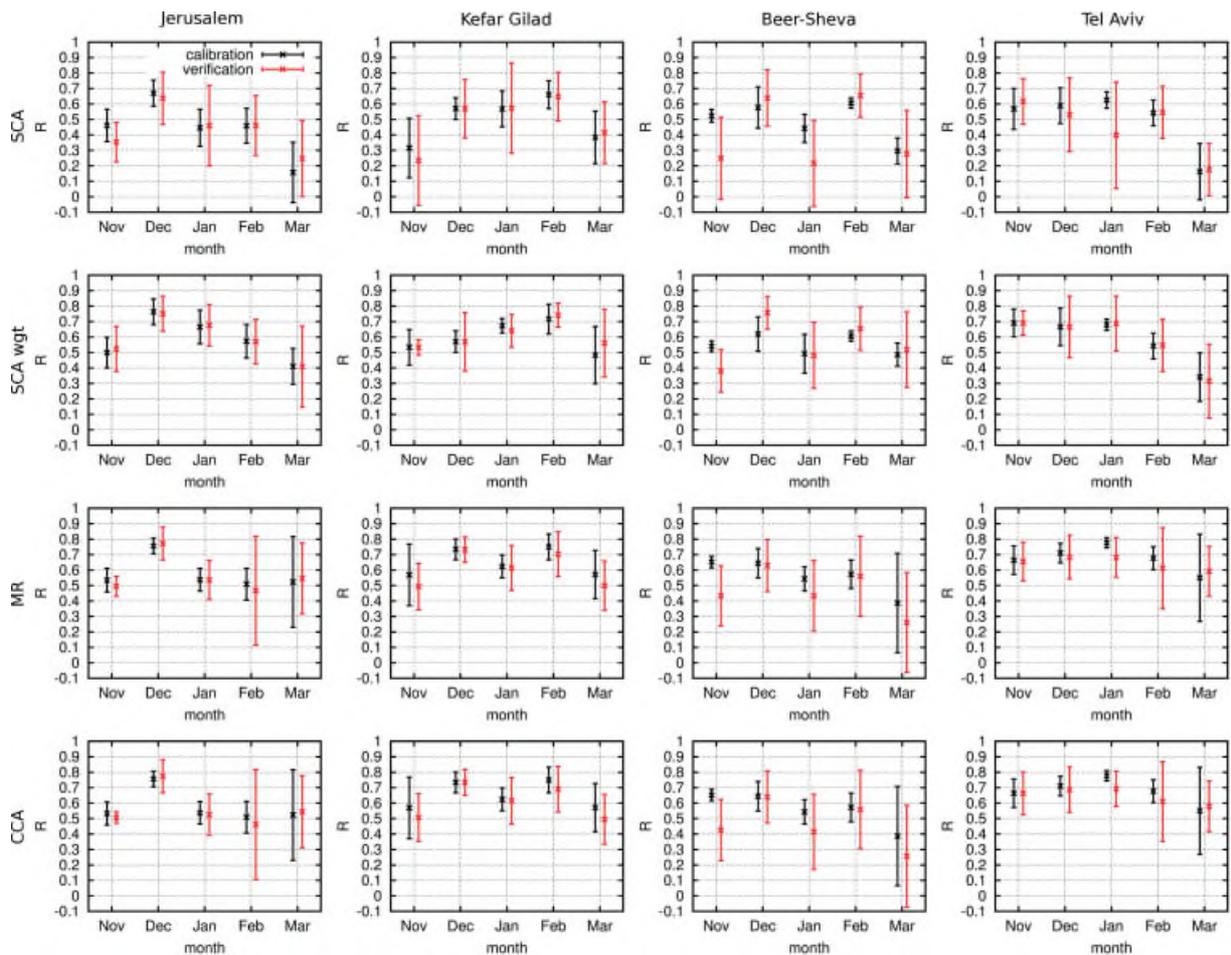


Figure 6. Calibration and verification ensemble mean correlation coefficients ( $R$ ) and standard deviations for all study months ( $x$ -axis) and all selected downscaling techniques: unweighted SANDRA cluster analysis (SCA), weighted SANDRA-CA (SCA wgt), multiple regression analysis (MR) and canonical correlation analysis (CCA). The only Predictor is sea level pressure. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

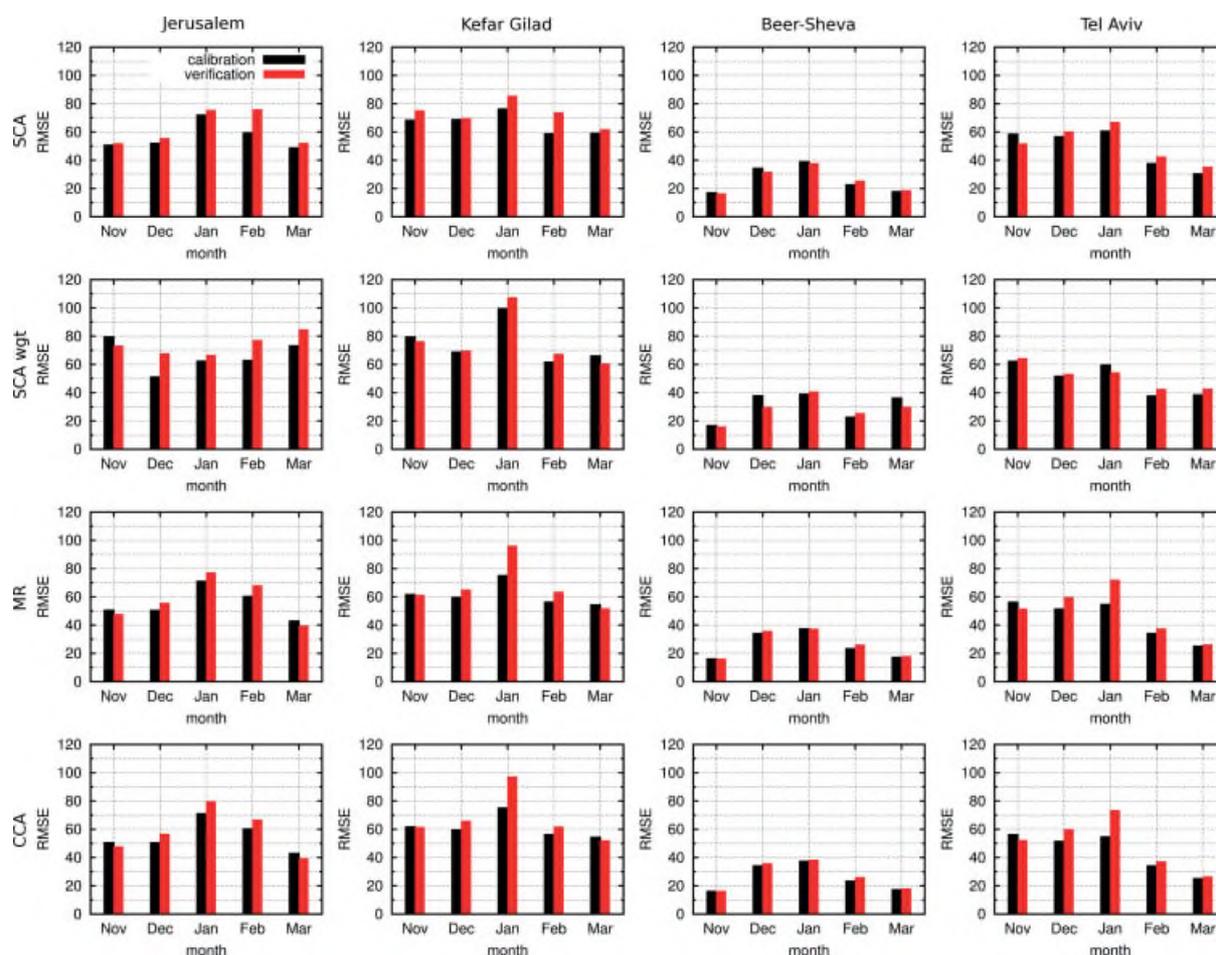


Figure 7. Calibration and verification ensemble mean root mean square errors (RMSE) for all study months ( $x$ -axis) and all selected downscaling techniques: unweighted SANDRA cluster analysis (SCA), weighted SANDRA-CA (SCA wgt), multiple regression analysis (MR) and canonical correlation analysis (CCA). The only predictor is sea level pressure. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

modelled rainfall amounts are squared emphasising these errors. However, RMSE remains a useful tool for comparison of several methods applied to a particular station and period. Regarding different techniques, in most cases weighted SCA yields the highest values (about 80 mm for January in Jerusalem and 100 mm in Kefar Gilad), whereas all other techniques achieve comparable RMSE values. In some cases for SCA, RMSE values are a little bit higher than those for MR and CCA.

Altogether, the regression-based methods CCA and MR show comparable performance regarding correlation coefficients and RMSE with marginal differences between individual cases. Both methods clearly outperform the classification technique with higher correlations and lower WEV with few exceptions (e.g. higher WEV in February and January for MR and CCA). SCA shows weak performance regarding both correlations and RMSE for most of the months of all stations. Weighting clearly increases explained variances but also increases RMSE. In some cases, weighted SCA even outperforms CCA and MR (e.g. Jan/Feb of Jerusalem), yielding higher correlation coefficients and comparable, or lower, RMSE values.

In most cases, SLP as a single predictor fails to explain precipitation on a local scale sufficiently. Model

performance is strongly dependent on month and station, therefore best values can be found for the wetter months and stations (winter months and Kefar Gilad/Tel Aviv). Both correlation coefficients and RMSE have to be considered carefully. Models yielding low correlations combined with low RMSE cannot describe the observed variability, even though deviations are small, whereas models achieving good correlations (and therefore, high percentages of explained variance  $R^2$ ) combined with high RMSE are capable to reproduce the observed variability but considerably under- or overestimate the observed precipitation amounts.

For further considerations, additional predictors are included in the models in order to evaluate how model performance can be improved.

#### 4.3. Multiple-predictor models

Including geopotential heights and relative vorticity of the 500-hPa level and moisture flux of the 1000-hPa level as additional predictors leads to more consistent results (Figures 8 and 9) and high agreements in correlation coefficients of the different techniques. RMSE values are substantially reduced. Beer-Sheva remains the station with lowest model performances, but compared

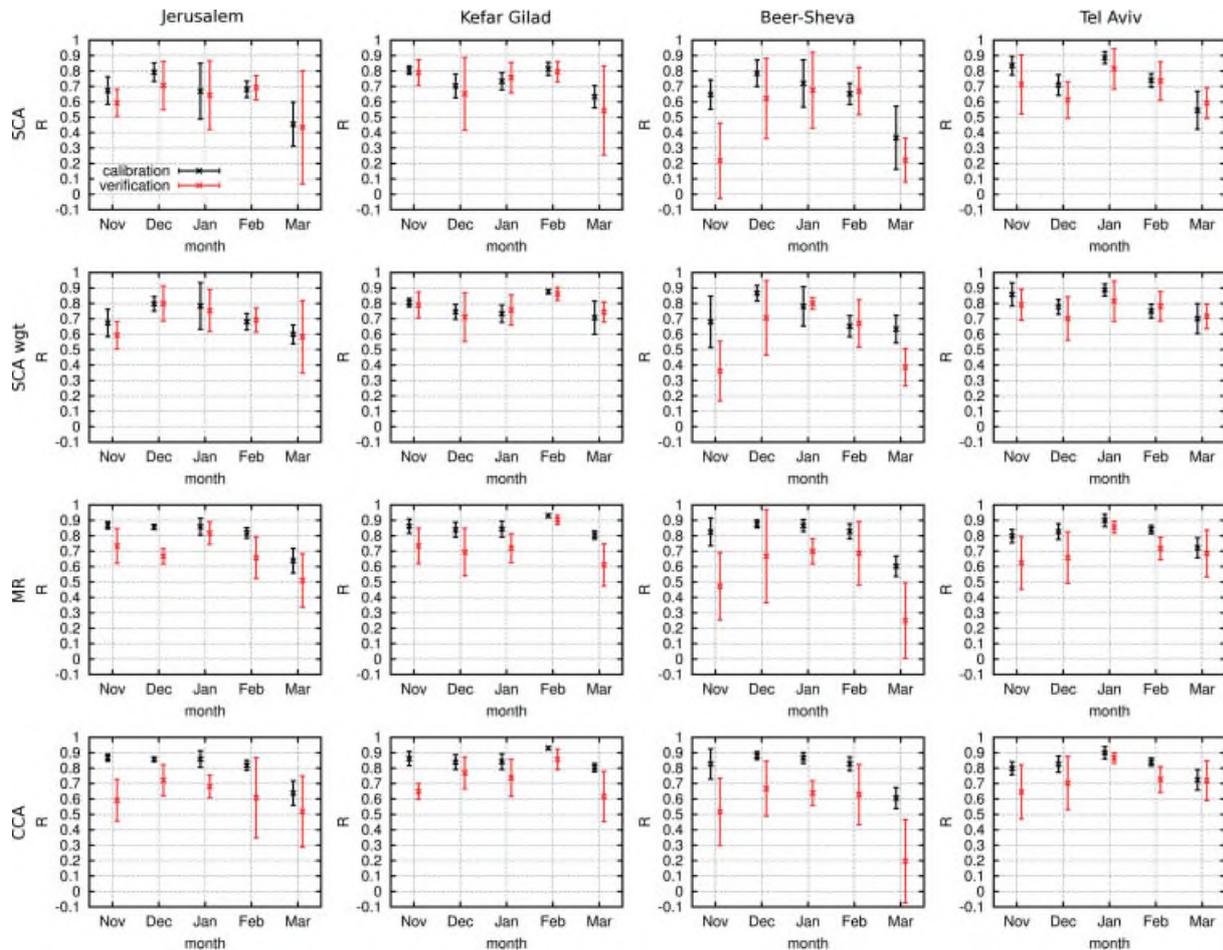


Figure 8. Calibration and verification ensemble mean correlation coefficients ( $R$ ) and standard deviations for all study months ( $x$ -axis) and all selected downscaling techniques: unweighted SANDRA cluster analysis (SCA), weighted SANDRA-CA (SCA wgt), multiple regression analysis (MR) and canonical correlation analysis (CCA). Predictors are sea level pressure, 500 hPa-geopotential heights, 500 hPa-vorticity, and 1000 hPa-moisture flux. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

to the SLP-models correlation coefficients are notably higher. Winter months'  $R$  values are mostly significant with CCA and MR performing comparably well and weighted SCA showing the best performance in most cases. Coefficients for the transition months are generally lower with higher variations within the ensembles. All other stations' modelled time series are found to be significantly correlated with the observed precipitation, except for unweighted SCA applied to Jerusalem and Kefar Gilad. RMSE values are comparable for all techniques but tend to be higher during verification for MR and CCA. Application of weighting leads to a considerable improvement, i.e. higher mean correlation coefficients and lower WEV. RMSE values are similar or marginally reduced. Altogether, CCA and MR again perform comparably well with little differences. For both techniques there are some cases (e.g. all months in Beer-Sheva, or November/December in Tel Aviv) with calibration correlation coefficients being higher than those for verification. Unweighted SCA performs better when including different predictors. In case of high WEV, weighting reduces the variations, leading to a result that is comparable to those of MR, also yielding lower RMSE values compared

to CCA. In some cases weighted SCA outperforms MR and CCA, especially as verification correlation coefficients do not drop.

Figure 10 shows the comparison of the four techniques' modelled monthly precipitation to the observed one, for one example of Jerusalem (Figure 10(a)) and one of Beer Sheva (Figure 10(b)). This selection aims to pick one successful example and one with low performance. For Jerusalem December precipitation of model period 2 (Section 3.3, Table II) was selected ( $R$ : 0.65–0.85, RMSE: 39–62 mm, weight = 12). Obviously, all models are capable of reproducing the long-term variance. Dependent on the applied technique, some peaks are over- or underestimated, such as 1971 or 1985. Figure 10(b) displays modelled and observed March precipitation for Beer-Sheva, model period 1 ( $R$ : 0.0–0.765, RMSE: 11–25 mm, weight = 9). None of the techniques succeeds in properly modelling precipitation.

To summarize findings in Sections 4.2 and 4.3, correlation coefficients and RMSE values were averaged over all ensemble members, months and stations. Figure 10 displays the model skills for the calibration (Figure 11(a) + (c)) and the verification periods (Figure 11(b) + (d)).

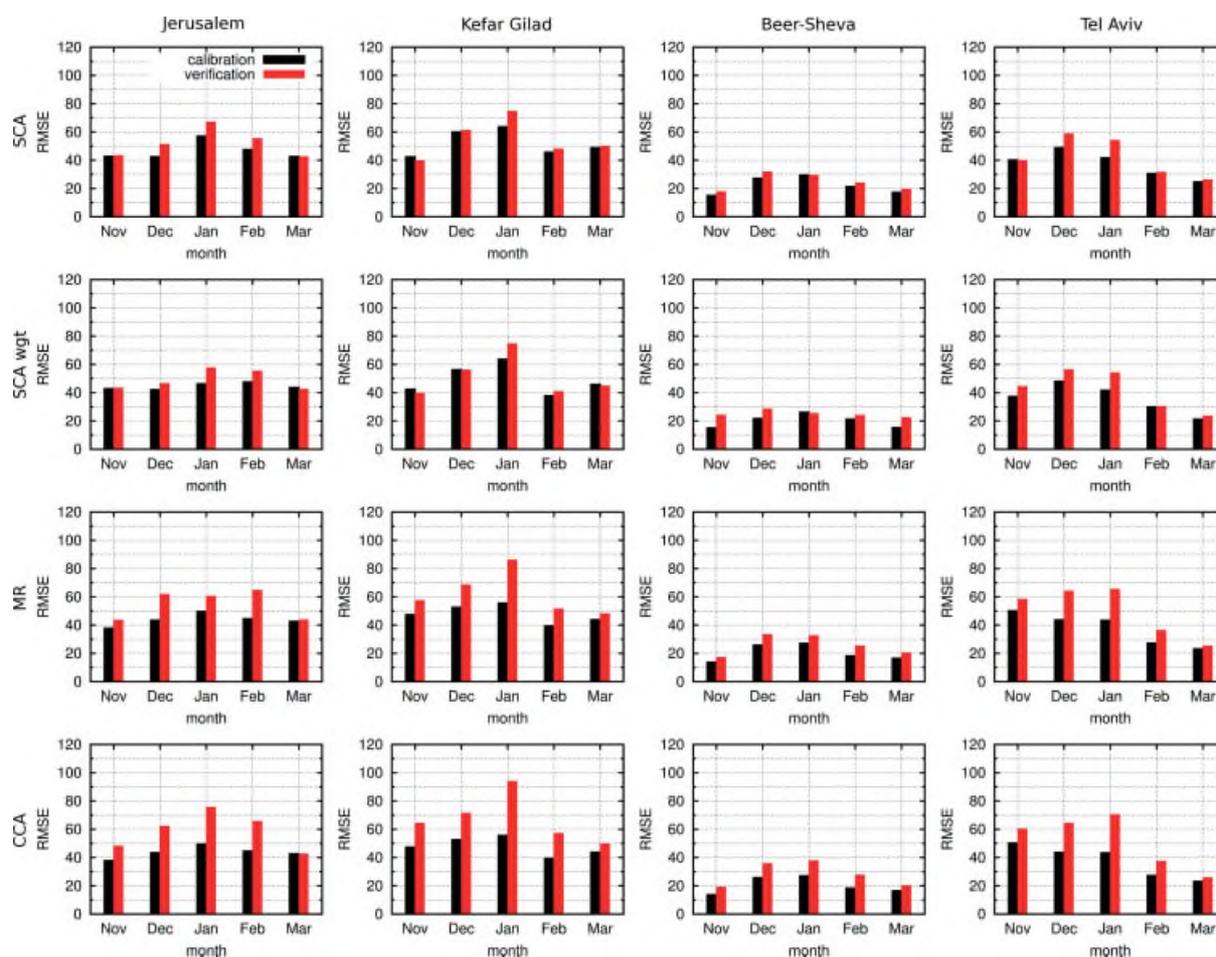


Figure 9. Calibration and verification ensemble mean root mean square errors (RMSE) for all study months ( $x$ -axis) and all selected downscaling techniques: unweighted SANDRA cluster analysis (SCA), weighted SANDRA-CA (SCA wgt), multiple regression analysis (MR) and canonical correlation analysis (CCA). Predictors are sea level pressure, 500 hPa-geopotential heights, 500 hPa-vorticity, and 1000 hPa-moisture flux. This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

Models based on SLP as single predictor yield average correlation coefficients of around 0.6 or lower with weakest results for unweighted SCA and weighted SCA performing comparably well to CCA and MR. RMSE values reach 50–60 mm with highest values for weighted SCA and lowest for CCA. Skills for calibration and verification differ only marginally. Additional predictors considerably improve results. Correlation coefficients reach values of 0.6 or higher and RMSE is reduced by 10–20 mm. For the calibration periods, MR and CCA yield the highest mean correlation 0.821/0.820, but during verification, they drop to 0.653/0.668, whereas, weighted SCA provides values of 0.749 and 0.701 for calibration and verification, respectively; the smaller difference between these two values indicates a higher model stability. Calibration RMSE values are comparable for the three techniques but higher during verification for MR and CCA. With respect to correlation coefficients and RMSE, the regression-based techniques can be considered as the best performing techniques for models based on the predictor SLP. For models based on multiple predictors, weighted SCA yields the best results, closely followed by MR and CCA.

When comparing all models and techniques it becomes clear that differences between techniques are smaller than those between the two predictor sets. Correlation coefficients are varying for one predictor set and different techniques only around 0.1, whereas, differences between both predictor sets for one particular technique are around 0.2.

#### 4.4. Synoptic comparisons

To study differences and similarities between objective unconditioned and precipitation-conditioned (weighted) SANDRA cluster analysis with respect to some dynamical characteristics, both classifications were compared to the semi-objective classification defined by Alpert *et al.* (2004). It includes 19 expert-defined classes belonging to 6 synoptic groups. One group – ‘winter lows’ – only contains Cyprus lows or lows located near Cyprus. Subdivision in this group (as for all others) was carried out according to the location and depth of the Cyprus lows (or other systems).

In this study, Cyprus lows are considered in particular as they are the main contributor to rainfall in Israel and Jordan (Goldreich, 2003; Saaroni, *et al.* 2010). According

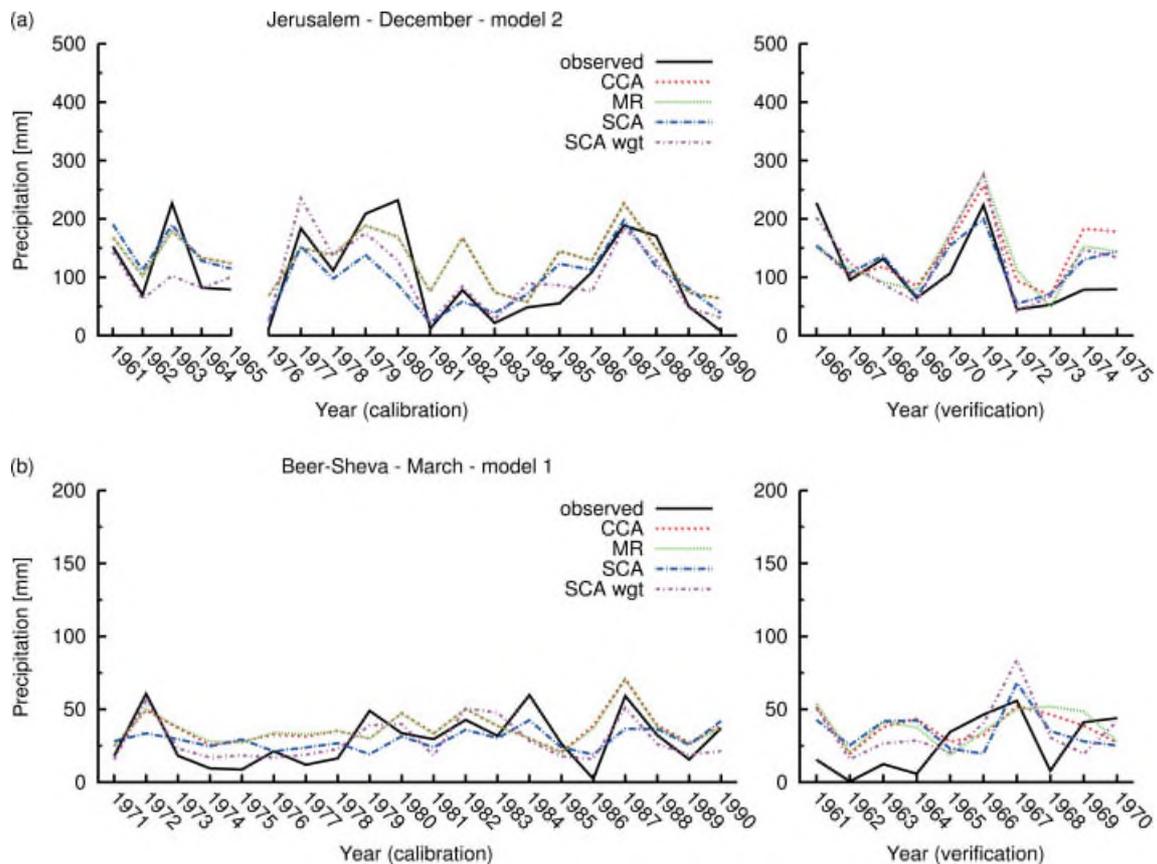


Figure 10. Observed and modelled monthly precipitation for Jerusalem (December, ensemble member m2, see Section 3.3, Table II) and Beer-Sheva (March, ensemble member m1, see Section 3.3, Table II), using different downscaling techniques: unweighted SANDRA cluster analysis (SCA), weighted SCA (SCA wgt), multiple regression analysis (MR) and canonical correlation analysis (CCA). This figure is available in colour online at [wileyonlinelibrary.com/journal/joc](http://wileyonlinelibrary.com/journal/joc)

to Saaroni *et al.* (2010) who use six of the seven winter low types or Cyprus lows, the latter are connected to different amounts of rainfall. For the example of January in Jerusalem (Figure 4 and Section 4.1), the occurrence of Cyprus low types is 72 and 68% for the unweighted wettest patterns #4 and #8. For pattern #4, mostly deep and shallow and lows to the East can be found which account for about 7–11 mm rainfall on average (average daily rainfall for different low types in Saaroni *et al.*, 2010). For pattern #8, the deep Cyprus low to the North which is connected to about 16 mm on average accounts for 38%. Other types accounting for less precipitation also occur. As rainfall occurrence of pattern #8 is lower (76%) than that of pattern #4 (88%), fewer days with more intense rainfall contribute to the pattern. For the residual percentages of rain days, other pressure patterns are found corresponding to the classification of Alpert *et al.* (2004). For patterns #3 and #4 of the weighted (weight = 5%) classification (which corresponds to unweighted SCA pattern #4), Cyprus low occurrences drop to 47 and 50% which agrees with a lower frequency of rain days and less mean precipitation. Pattern #9 (which corresponds to unweighted SCA pattern #8), which yields the highest average amount of rainfall (44.2 mm) with a rain-day frequency of 100%, is connected to a Cyprus low

occurrence of 96%, with 35% deep lows to the North, and 38% lows to the East. Rainfall on the remaining 27% of days is connected to other synoptic types. The number of 9 clusters appears to be insufficient to split up this pattern into different synoptic types connected to days with high precipitation amounts. However, a pre-defined larger number of clusters would lead to a classification consisting of patterns with very specific conditions and low pattern occurrences. As a classification of that kind cannot be transferred to independent periods, some within-pattern variability has to be accepted.

## 5. Summary and conclusions

To contribute to the improvement of regional climate projections, this study compared the performance of several statistical downscaling techniques for station-based precipitation in the ME region. Two classical techniques, MR and CCA, as well as a novel approach, unconditioned and precipitation-conditioned (weighted) SCA, were evaluated for the main rainy season (November–March) using cross-validation. An ensemble of five models was established for each of the selected months and validated using two different skill scores, correlation coefficients ( $R$ ) between observed and modelled precipitation and RMSE, as well as the WEV. Prior to the comparison of the

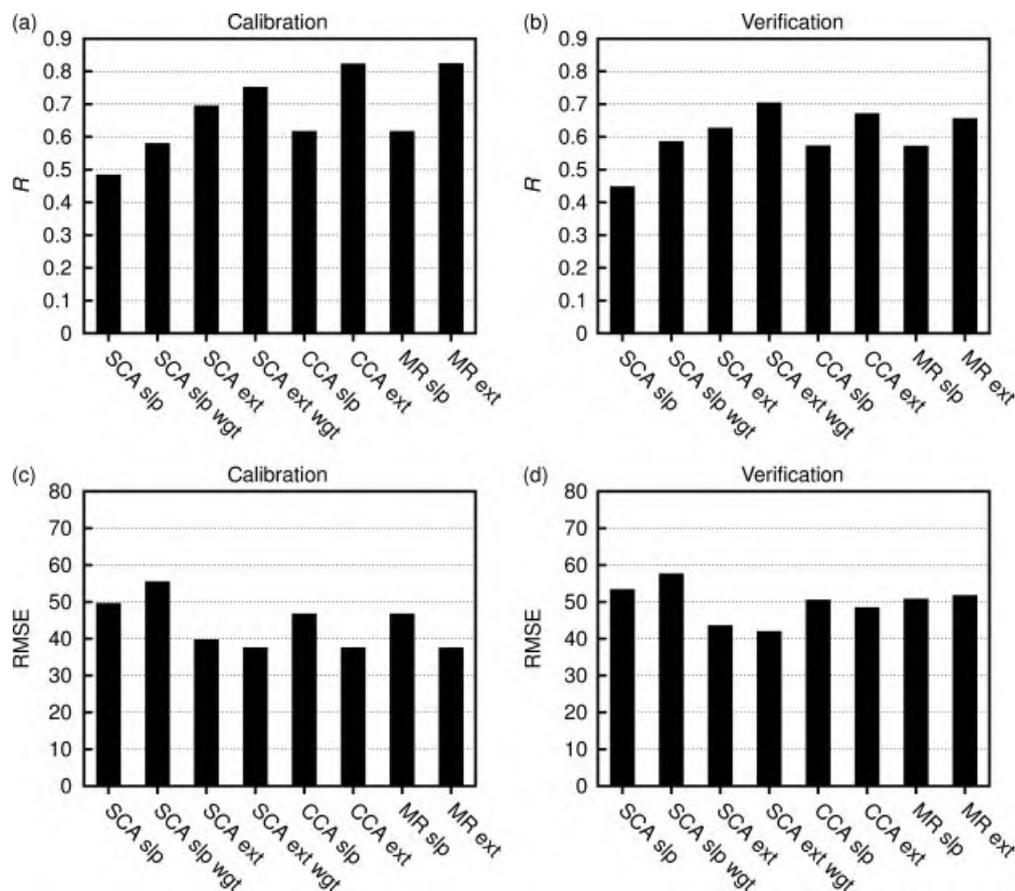


Figure 11. correlation coefficients ( $R$ , parts (a) and (b)) and root mean square errors (RMSE, parts c and d) for calibration (parts (a) and (c)) and verification periods (parts (b) and (d)) averaged over all ensemble members, months and stations, for different downscaling techniques (x-axis) and predictor combinations (indicated as slp for SLP-models or ext for multiple predictor-models): unweighted SANDRA cluster analysis (SCA), weighted SCA (SCA wgt), multiple regression analysis (MR) and canonical correlation analysis (CCA).

different techniques, the application of different weights to the precipitation-conditioned SCA was evaluated. The best results were selected and applied for the comparison to the remaining techniques. To evaluate the effect of selected predictor variables on the modelling results, two different predictor sets were used, one set consisting of SLP as the only predictor, and another set using further atmospheric predictors: geopotential heights, relative vorticity, and  $MF_u$ .

The main conclusions that can be drawn from the analyses can be summarized as follows:

- All selected downscaling techniques were successfully applied for the study region
- For the synoptic downscaling technique, the inclusion of precipitation into the clustering leading to a precipitation-conditioned (weighted) cluster analysis considerably improves model performance with results that are comparable to the regression-based methods. SLP-based unweighted SCA that does not include information on precipitation, produces circulation patterns that do not represent a particular weather state or rainfall amount, but include days with highly variable rainfall amounts. Variability within such a pattern is higher than variability across different patterns. Weighted SCA, however, is capable to find

patterns that aggregate days with similar precipitation amounts, reducing within-pattern variability for the benefit of more dissimilar circulation patterns with different precipitation characteristics. This also affects downscaling in a favourable way: Stability of the circulation-precipitation relationships and downscaling model performances are distinctly increased

- SLP as single predictor does not perform sufficiently well. Resulting downscaling model performance is strongly dependent on the selected month, station or technique
- The inclusion of  $HGT$ , relative vorticity and  $MF_u$  as additional predictors considerably improves model performance, regardless of the particular technique used. Correlation coefficients are higher, RMSEs are reduced, and a lower within-ensemble variance indicates enhanced model stability
- MR and CCA yield the best results when only SLP is used as predictor compared to the remaining techniques. Unweighted SCA shows lowest performance. Weighting improves correlation coefficients but also increases RMSE
- SCA performs best when further predictors are included. It is closely followed by MR and CCA. Even unweighted SCA performs reasonably well with

additional predictors, indicating that especially the inclusion of humidity information of some kind (i.e.  $MF_u$  in this case) results in a weather pattern classification with dissimilar patterns accounting for different rainfall amounts. Weighting further increases dissimilarity in some cases. Thus, weighted multiple-predictor SCA proves to be a well performing downscaling technique besides CCA and MR; it might have the potential to outperform the regression-based techniques.

- When considering all models from different predictor sets and techniques, results clearly show that the differences between techniques are smaller than those between the two predictor sets. Therefore, a careful choice of predictors is very important before the best technique can be identified.

Considering water scarcity and the associated risk for social and political conflicts in the eastern Mediterranean/Middle East region, further downscaling studies for recent and future climates as well as impact studies are required. Future work should apply the described techniques and analyse, e.g. the impact of future climate change on water availability.

## Acknowledgements

The authors acknowledge financial support and data provided by the GLOWA Jordan River Project which is funded by the German Federal Ministry of Education and Research (BMBF). The authors were further supported by the European Commission under contract EVK2-CT-2002-00161 EMULATE.

## References

- Alpert P, Getenio B, Seter I, Dekker D, Zak-Rosenthal R. 1987. *Wind energy potential over complex terrain with a one-level mesometeorological model and summer classification of synoptic situations in Israel*. Department of Energy and Infrastructure and Israeli Meteorological Service, Israel.
- Alpert P, Neeman B, Shay-El Y. 1990. Climatological analysis of Mediterranean cyclones using ECMWF data. *Tellus* **42A**: 65–77.
- Alpert P, Osetinsky I, Ziv B, Shafir H. 2004. Semi-objective classification for daily synoptic systems, application to the Eastern Mediterranean climate change. *International Journal of Climatology* **24**: 1013–1011, DOI:10.1002/joc.1036.
- Anagnostopoulou C, Flocas H, Maheras P, Patrikas I. 2004. Relationship between atmospheric circulation types over Greece and Western-Central Europe during the period 1958–97. *International Journal of Climatology* **24**: 1745–1758.
- Bárdossy A, Stehlik J, Caspary HJ. 2002. Automated objective classification of daily circulation patterns for precipitation and temperature downscaling based on optimized fuzzy rules. *Climate Research* **22**: 11–22.
- Barnett T, Preisendorfer R. 1987. Origins and levels of monthly and seasonal forecast skill for the United States surface air temperatures determined by canonical correlation analysis. *Monthly Weather Review* **115**: 1825–1850.
- Beck C, Philipp A. 2010. Evaluation and comparison of circulation type classifications for the European domain. *Physics and Chemistry of the Earth* **35**: 374–387.
- Buishand TA, Brandsma T. 1997. Comparison of circulation classification schemes for predicting temperature and precipitation in the Netherlands. *International Journal of Climatology* **17**: 875–889.
- Buishand TA, Shabalova MV, Brandsma T. 2004. On the Choice of the Temporal Aggregation Level for Statistical Downscaling of Precipitation. *Journal of Climate* **17**: 1816–1827.
- Busuioc A, Tomozeiu R, Cacciamani C. 2008. Statistical downscaling model based on canonical correlation analysis for winter extreme precipitation events in the Emilia-Romagna region. *International Journal of Climatology* **28**: 449–464, DOI:10.1002/joc.1547.
- Easterling DR. 1999. Development of regional climate scenarios using a downscaling approach. *Climatic Change* **41**: 615–634.
- Fereday DR, Knight JR, Scaife AA, Folland CK, Philipp A. 2008. Cluster analysis of North Atlantic/European circulation types and links with tropical Pacific sea surface temperatures. *Journal of Climate* **21**(15): 3687–3703, DOI:10.1175/2007JCLI1875.1.
- Goldreich Y. 2003. *The Climate of Israel: Observation, Research and Application*. Kluwer: Academic/Plenum Publishers, New York.
- Goodess CM, Palutikof JP. 1998. Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling. *International Journal of Climatology* **10**: 1051–1083.
- Giorgi F, Bi X, Pal JS. 2004a. Mean, interannual variability and trends in a regional climate change experiment over Europe. Part I: Present day climate (1961–1990). *Climate Dynamics* **22**: 733–756.
- Giorgi F, Bi X, Pal JS. 2004b. Mean, interannual variability and trends in a regional climate change experiment over Europe. Part II: Future climate scenarios (2071–2100). *Climate Dynamics* **23**: 839–858.
- Giorgi F, Hewitson B, Christensen J, Hulme M, von Storch H, Whetton P, Jones R, Mearns L and Fu C. 2001. Regional Climate Information – Evaluation and projections. In *Climate Change 2001: The Physical Science Basis*. Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (eds). Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA; pp. 583–638.
- Haylock MR, Cawley GC, Harpham C, Wilby RL, Goodess CM. 2006. Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology* **22**: 1397–1415.
- Hertig E, Jacobeit J. 2008. Assessments of Mediterranean precipitation changes for the 21st century using statistical downscaling techniques. *International Journal of Climatology* **28**: 1025–1045, DOI:10.1002/joc.1597.
- Huth R. 2002. Statistical Downscaling of Daily Temperature in Central Europe. *Journal of Climate* **15**: 1731–1742.
- Huth R. 2004. Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors. *Journal of Climate* **17**: 640–652, DOI:10.1175/1520-0442(2004)017.
- Huth R, Beck C, Philipp A, Demuzere M, Ustrnul Z, Cahynová M, Kyselý J, Tveito OE. 2008. Classifications of Atmospheric Circulation Patterns – Recent Advances and Applications. *Trends and Directions in Climate Research: Ann. N.Y. Acad. Sci.* **1146**: 105–152, DOI:10.1196/annals.
- Huth R, Kyselý J, Dubrovsky M. 2001. Time structure of observed GCM-simulated downscaled and stochastically generated daily temperature series. *Journal of Climate* **14**: 4047–4061.
- IPCC. 2007. *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA; p. 996.
- Jacobeit J. 1993. Regionale Unterschiede im atmosphärischen Zirkulationsgeschehen bei globalen Klimaveränderungen. *Die Erde* **124**: 63–77.
- Jacobeit J. 1996. Atmospheric Circulation Changes Due to Increased Greenhouse Warming and its Impact on Seasonal Rainfall in the Mediterranean Area. – In *Climate Variability and Climate Change Vulnerability and Adaptation*. Nemesova I (ed), Institute of Atmospheric Physics, Czech Academy of Sciences, Praha, 71–80.
- Jacobeit J. 2010. Classifications in Climate Research. *Physics and Chemistry of the Earth* **35**: 411–421.
- Jones PD, Hulme M, Briffa KR. 1993. A comparison of Lamb Circulation Types with an objective classification derived from grid-point mean-sea-level pressure data. *International Journal of Climatology* **13**: 655–663.
- Kahana R, Ziv B, Enzel Y, Dayan U. 2002. Synoptic climatology of major floods in the Negev desert, Israel. *International Journal of Climatology* **22**: 867–882.
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Ebisduzaki W, Higgins W, Janowiak J, Mo K, 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.

- Kioutsioukis I, Melas D, Zanis P. 2008. Statistical downscaling of daily precipitation over Greece. *International Journal of Climatology* **28**: 679–691, DOI:10.1002/joc.1557.
- Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuziak W, Kanamitsu M, Kousky V, van den Dool H, Jenne R, Fiorino M. 2001. The NCEP/NCAR 40-year reanalysis project: Monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society* **82**: 247–268.
- Koplowitz R. 1973. *An objective classification of synoptic pressure field patterns of the Eastern Mediterranean basin for use in climatological studies*. MSc thesis, Hebrew University, Jerusalem.
- Kunstmann H, Suppan P, Heckl A, Rimmer A. 2007. Regional climate change in the Middle East and impact on hydrology in the Upper Jordan catchment. In *Quantification and reduction of predictive uncertainty for sustainable water resources management. Proceedings of Symposium HS2004 at IUGG2007*, Perugia, July 2007, Publication 313; pp. 141–149.
- Kutiel H, Paz S. 1998. Sea level pressure departures in the Mediterranean and their relationship with monthly rainfall conditions in Israel. *Theoretical and Applied Climatology* **60**: 93–109.
- Laux P, Heckl A, Jacobeit J, Bárdossy A, Kunstmann H. 2008b. Linking droughty and wet weather in the Jordan catchment with atmospheric circulation patterns. In *Cambio climático regional y sus impactos*. Sigró J, Brunet M, Aguilar E (eds). *Publicaciones de la Asociación Española de Climatología (AEC) series A*, no. 6, pp. 375–386.
- Laux P, Kunstmann H, Lutz K, Jacobeit J, Beck C, Philipp A, Bárdossy A. 2008a. Comparison of different objective atmospheric circulation pattern analyses in the Jordan region. IAHS Publication, Prague, Czech Republic, 15–18 September 2008 (CD-Rom).
- Li L, Bozec A, Somot S, Béranger K, Bouruet-Aubertot P, Sevault F, Crépon M. 2006. Regional Atmospheric, Marine Processes and Climate Modelling. In *Mediterranean climate variability and predictability*. Lionello P (ed). Elsevier: Amsterdam; pp. 373–397.
- Maheras P, Flocas HA, Patrikas I, Anagnostopoulou C. 2001. A 40 year objective climatology of surface cyclones in the Mediterranean region: spatial and temporal distribution. *International Journal of Climatology* **21**: 109–130.
- Michaelson J. 1987. Cross-Validation in Statistical Climate Forecast Models. *Journal of Climate and Applied Meteorology* **26**: 1589–1600.
- Murphy J. 2000. Predictions of climate change over Europe using statistical and dynamical downscaling techniques. *International Journal of Climatology* **20**: 489–501.
- Philipp A, Bartholy J, Beck C, Erpicum M, Esteban P, Huth R, James P, Jourdain S, Krennert T, Lykoudis S, Michaelides S, Pianko K, Post P, Rassilla Álvarez D, Spekat A, Tymvios FS. 2010. COST733CAT – a database of weather and circulation type classifications. *Physics and Chemistry of the Earth* **35**: 360–373.
- Philipp A, Della-Marta P, Jacobeit J, Fereday DR, Jones PD, Moberg A, Wanner H. 2007. Long-term variability of daily North Atlantic-European pressure patterns since 1850 classified by simulated annealing. *Journal of Climate* **20**: 4066–4095.
- Preisendorfer RW. 1988. *Principal Component Analysis in Meteorology and Oceanography. Developments in Atmospheric Science* 17. Elsevier: Amsterdam.
- Ribera P, García R, Díaz HF, Gimeno L, Hernández E. 2000. Trends and interannual oscillations in the main sea-level pressure patterns over the Mediterranean, 1955–1990. *Geophysical Research Letters* **27**: 1143–1146.
- Ronberg B. 1984. *An objective weather typing system for Israel: a synoptic climatological study*. PhD thesis, Hebrew University, Jerusalem.
- Saaroni H, Halfon N, Ziv B, Alpert P, Kutiel H. 2010. Links between the rainfall regime in Israel and location and intensity of Cyprus lows. *International Journal of Climatology* **30**: 1014–1025.
- Shafir H, Ziv B, Neeman BU, Alpert P. 1994. Objective classification of synoptic situations and implication to economical wind energy availability in Israel. *Journal of Israel Meteorological Society* **3**: 29–49.
- Stehlík J, Bárdossy A. 2003. Statistical comparison of European circulation patterns and development of a continental scale classification. *Theoretical and Applied Climatology* **76**: 31–46.
- Suppan P, Kunstmann H, Heckl A, Rimmer A. 2008. Impact of Climate Change on Water Availability in the Near East. In *Climatic Changes and Water Resources in the Middle East and in North Africa*. Zereini F, Hötzl H (eds). Springer-Verlag: Berlin, Heidelberg, New York.
- von Storch H, Zwiers FW. 1999. *Statistical Analysis of Climate Research*. Cambridge University Press: Cambridge.
- Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO. 2004. *Guidelines for use of Climate Scenarios Developed from Statistical Downscaling Methods*. IPCC Task Group on Data and Scenario Support for Impact and Climate Analysis. www.ipcc-data.org/guidelines/dgm.no2.v1.09.2004.pdf (14.10.2009).
- 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 Research* **34**: 2995–3008.
- Wilks D. 2005. *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press Inc.: USA.
- Zangvil A, Karas S, Sasson A. 2003. Connection between eastern Mediterranean seasonal mean 500 HPA height and sea-level pressure patterns and the spatial rainfall distribution over Israel *International Journal of Climatology* **23**: 1567–1576.
- Ziv B, Saaroni H, Alpert P. 2006. The summer climate in Israel and its connection to the Monsoon. MERCHAVIM special issue in honour of Prof. Arie Bitan, 315–331.
- Zorita E, von Storch H. 1999. Analog method as a simple statistical downscaling technique: comparison with more complicated methods. *International Journal of Climatology* **12**: 2474–2489.