

Influence of similarity measures on the performance of the analog method for downscaling daily precipitation

C. Matulla · X. Zhang · X. L. Wang · J. Wang ·
E. Zorita · S. Wagner · H. von Storch

Received: 2 March 2007 / Accepted: 22 May 2007
© Springer-Verlag 2007

Abstract This study examines the performance of the analog method for downscaling daily precipitation. The evaluation is performed for (1) a number of similarity measures for searching analogs, (2) various ways to include the past atmospheric evolution, and (3) different truncations in EOF space. It is carried out for two regions with complex topographic structures, and with distinct climatic characteristics, namely, California's Central Valley (together with the Sierra Nevada) and the European Alps. NCEP/NCAR reanalysis data are used to represent the large scale state of the atmosphere over the regions. The assessment is based on simulating daily precipitation for 103 stations for the month of January, for the years 1950–2004 in the California region, and for 70 stations in the European Alps (January 1948–2004). Generally, simulated precipitation is in better agreement with observations in the California region than in the European Alps. Similarity measures such as the Euclidean norm, the sum of absolute differences and the angle between two atmospheric states perform better than measures which introduce additional weightings to principal components (e.g., the Mahalanobis distance). The best choice seems dependent upon the target variable. Lengths of wet spells, for instance, are best simulated by using the angular similarity measure. Overall, the Euclidean norm performs satisfactorily in most cases and hence is a reasonable first choice, whereas the use of Mahalanobis dis-

tance is less advisable. The performance of the analog method improves by including large-scale information for bygone days, particularly, for the simulation of wet and dry spells. Optimal performance is obtained when about 85–90% of the total predictor variability is retained.

1 Introduction

The analog method (AM) has been commonly used in weather prediction (Elliot 1951; Baur 1951; Lorenz 1969) and in seasonal forecasting (Livezey and Barnston 1988). Lorenz (1969) studied atmospheric predictability by the use of analogs but achieved poor results as the analogs were drawn from sparse weather archives (van den Dool 1994). Livezey et al. (1994) compared analog prediction systems, developed in the US (Livezey and Barnston 1988) and the former Soviet Union (Gruza and Ran'kova 1986) for the purpose of enhancing US seasonal temperature predictions. Even though seasonal forecasts are presently processed by ensemble forecasts of a number of general circulation models (GCMs), an analog prediction scheme still can act as a benchmark. Recently, Hamill and Whitaker (2006) explored a set of analog techniques for the statistical correction of weather forecasts. A successful application of AM requires an extensive archive of observations, and depends on the size and complexity of the considered region. In general it is more difficult to find a close match for an atmospheric state over a large region of complex structure than over a small, simply structured region.

Zorita et al. (1995) and Zorita and von Storch (1999) introduced AM into the field of downscaling. One objective

C. Matulla (✉) · X. Zhang · X. L. Wang ·
J. Wang
Climate Research Division, Environment Canada,
4905 Dufferin Street, Toronto, ON, Canada
e-mail: Christoph.Matulla@ec.gc.ca

E. Zorita · S. Wagner · H. von Storch
Institute for Coastal Research, GKSS,
Geesthacht, Germany

of downscaling is to generate statistics of local scale climate features that are consistent with the large scale atmospheric state (von Storch et al. 1993). When downscaling is used to generate future climate on local scales, archives should ideally cover the range of all possible future states.

The AM is a simple downscaling method, easy to implement and featuring two merits: AM uses observed weather patterns and hence, the spatial covariance structure of local scale weather is maintained in the simulated fields, important for hydrological studies, for instance, that rely on spatially meaningful patterns and; AM does not assume the form of probability distribution of downscaled variables, making it easy to construct scenarios for non-normally distributed variables such as daily precipitation.

AM is based on the selection of similar atmospheric states, and hence, its performance is dependent upon how similarity is quantified. Therefore, downscaling skill depends on the specific similarity measure used, and more generally on the overall selection process. Zorita and von Storch (1999) used an empirical orthogonal function (EOF) analysis to filter the noise on spatial small scales and to reduce the dimensionality of daily atmospheric states (see e.g., von Storch and Zwiers 1999). Subsequently they used the Euclidean norm for principal components to identify closest atmospheric states. They also suggested the use of different weights of the principal components. Toth (1991) compared nine different similarity measures for forecasting circulation at 700-hPa height without filtering the atmospheric data. He found that the mean absolute difference in the gradient of the height results in best circulation forecasts and concluded in favor of root-mean-square difference over correlation as a similarity measure.

The main purpose of the present study is to examine the dependence of AM's performance on the way similarity between large scale patterns is quantified. AM is evaluated for the regions of California and the European Alps, both with complex terrain but with different precipitation climates. Like any other statistical downscaling method, the success of AM relies on the existence of a strong relation between predictors and predictand. Therefore, we focus on January, as winter is when the large scale state of the atmosphere exerts strongest influence on precipitation in these regions. Based on this condition AM is expected to feature reasonable performance, a necessary prerequisite to study the sensitivity of the results to different similarity measures. Summer precipitation characteristics are certainly of great importance too. Especially changes in dry spells play a key role in impact considerations. However, their investigation is beyond the scope of the present paper, because during summer the link between the scales is rather loose (see e.g., Hamill and Whitaker 2006).

2 Study regions and data

2.1 Study regions

The Sierra Nevada, east of California's Central Valley, is a meridional oriented mountain ridge extending over 600 km. It is located about 250 km inland from the Pacific ocean and the highest peaks exceed 4,000 m. Altitude increases from Mt. Lassen ($\approx 3,200$ m) southward to Mt. Whitney ($\approx 4,400$ m). South-westerlies in January are the main source of precipitation for California. Compared to summer, the westerlies in winter are shifted southwards advecting humid air from the Pacific Ocean into California, that together with orographic lifting along the Sierra Nevada, makes winter the major precipitation season (Pandey et al. 1999). ENSO exerts a strong influence on precipitation as well (Ropelewski and Halpert 1986) causing above average precipitation during La Niña years and below average totals during El Niño years (Schonher and Nicholson 1989). Mean precipitation is largest in the north at high altitudes and lowest in the south at low elevations. Precipitation totals generally increase with altitude from West to East, peak at the summits of the Sierra Nevada and rapidly decrease further eastward (see Fig. 1, left column). Hence, the Sierra Nevada acts as a meteorological divide supplying water for agriculture and California's cities. The California region (hereafter CA) spans several climatic zones ranging from Desert Climate/Hot Steppe at low elevations (<150 m), to Mediterranean Climate at higher elevations (150–750 m) to Alpine Climate in the mountains (>750 m).

Contrary to the Sierra Nevada, the European Alps (hereafter EA) extent over the study region parallel to the westerlies. This region in Central Europe is about 1,000 km away from the Atlantic Ocean. There are areas of low altitude (<100 m), mountain peaks close to 3,800 m, and steep valleys. The region is under the influence of three air masses originating in the Atlantic, the Mediterranean and the European continent. The Alpine ridge acts as a barrier to these air masses. In January, the northern Alps receive most precipitation as a result of advection from the northern North Atlantic and the North Sea. The eastern part of the region is influenced by dry and cold continental air. Areas in the south of the main Alpine ridge receive more precipitation than the eastern parts. Largest amounts of precipitation occur north and north-west of the Alpine ridge (see Fig. 1 right column). Inner-Alpine valleys are particular dry due to the shadowing effect of the northern and southern ranges. Differing from the California region, winter is the driest season in the investigated part of the European Alps. Auer et al. (2001) provides a detailed overview of the climate of this region.

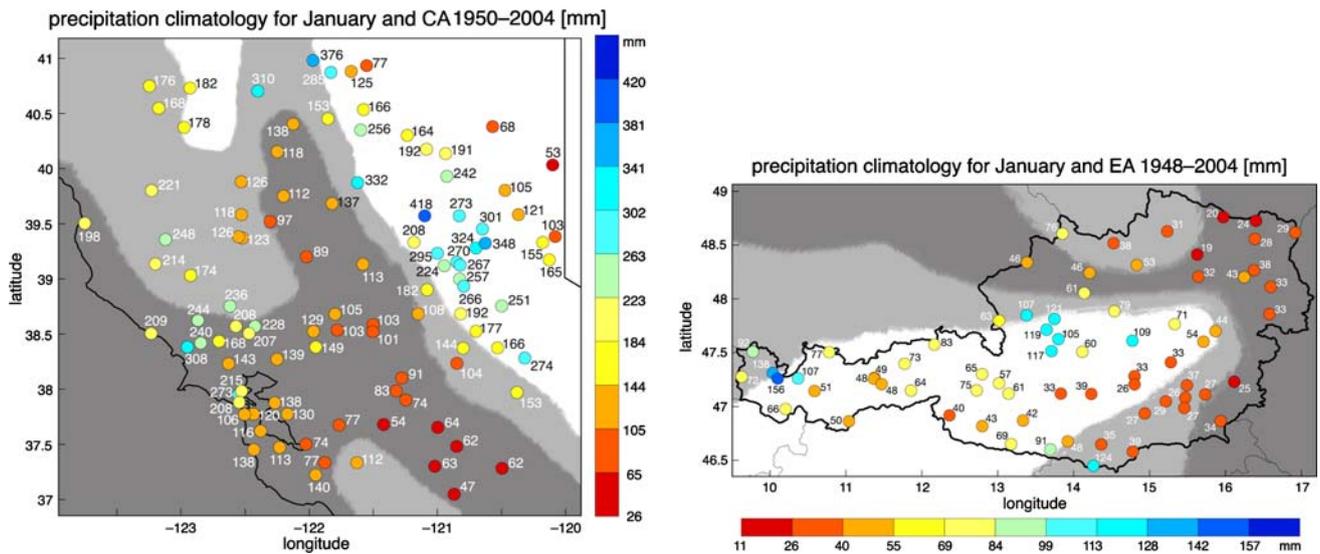


Fig. 1 Long term mean January precipitation at stations in the Californian region (*left*) and in the European Alps (*right*). Totals (mm) are listed next to the stations. *Colorbars* indicate the range of

precipitation amount (note the different ranges). *Grey shadings* characterize altitudinal belts (increasing from *dark-grey* to *white*) referred to in the text

2.2 Datasets

The first step in the analysis involves the selection of large scale predictors that are closely linked to local scale precipitation. There is no universal rule which fields to choose, but a number of downscaling studies for precipitation suggest the use of an atmospheric circulation and/or a moisture field in the lower troposphere (see e.g., IPCC 2001 for a list of studies). Here we use sea level pressure and specific humidity at 700 hPa to represent the atmospheric circulation and moisture, respectively.

Daily fields of mean sea level pressure (SLP) and specific humidity at 700 hPa (SH7) are retrieved from the NCEP/NCAR reanalysis data archives (Kalnay et al. 1996). Datasets are provided on a $2.5^\circ \times 2.5^\circ$ lat-long grid. Geographical sectors cover $20^\circ\text{N}/150^\circ\text{W}$ to $50^\circ\text{N}/110^\circ\text{W}$ and $35^\circ\text{N}/10^\circ\text{W}$ to $65^\circ\text{N}/30^\circ\text{E}$ for CA and EA, respectively. Daily precipitation levels from 103 stations in California drawn from the GHCN dataset (Gleason 2002), covering 1950–2004, (assembled by the National Climatic Data Center (NCDC), NOAA) are employed in the analysis. In the European Alps, daily precipitation levels for 70 stations are employed in the analysis, covering the period 1948–2004 (data are provided by the Austrian weather service, see Schöner et al. 2003).

3 Method

Subsection 3.1 lists several similarity measures acting between two large scale atmospheric states that are used in this study. These measures are the basis for the analog-selection process, outlined in Subject. 3.2.

3.1 Similarity measures

To reduce the dimensionality and to filter the atmospheric data, anomalies are projected onto the leading EOFs. To rescale the SLP and SH7 fields we divide them by their averaged standard deviations. Rescaling can be done using the square root of the averaged variances as weights as well, but differences are small in our case. Similarity measures are then computed using the values of the leading principal components (PCs) of the joint daily SLP and SH7 anomaly fields. The sensitivity to EOF-truncation is examined by varying the number of PCs from 5 to 26 in steps of 3. Five measures are considered to quantify similarity between atmospheric states. Let \mathbf{x} and \mathbf{y} be vectors containing the values of the leading n PCs and \mathbf{z} be their difference ($\mathbf{z} = \mathbf{x} - \mathbf{y}$). We consider the following measures of similarity:

EUK: L_2 norm of \mathbf{z} , the Euclidean distance between \mathbf{x} and \mathbf{y} :

$$\|\mathbf{z}\|_2 = \left(\sum_{i=1}^n z_i^2 \right)^{\frac{1}{2}} \tag{1}$$

XPL: Weighting the components with the corresponding explained variance (p_i , see e.g., Zorita and von Storch 1999):

$$\|\mathbf{z}\|_w = \left(\sum_{i=1}^n p_i z_i^2 \right)^{\frac{1}{2}} \tag{2}$$

SUM: The sum of absolute differences in the components:

$$\|\mathbf{z}\|_1 = \sum_{i=1}^n |z_i| \tag{3}$$

Cos: The negative cosine of the angle between \mathbf{x} and \mathbf{y} :

$$-\cos(\angle(\mathbf{x}, \mathbf{y})) = -\frac{\sum_{i=1}^n x_i y_i}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} \tag{4}$$

MAH: Mahalanobis distance as taken from Yambor et al. (2002), where λ_i is the i th eigenvalue corresponding to the i th component:

$$d(\mathbf{x}, \mathbf{y}) = -\sum_{i=1}^n \frac{x_i y_i}{\sqrt{\lambda_i}} \tag{5}$$

Abbreviations introduced above will be used hereafter. The measures evaluate similarity between state-vectors by different mappings of the components. EUK and SUM do not introduce component sensitive weights, but map rather differently the way differences are distributed across the components. XPL weights the components by the explained variance and MAH divides them by the square root of the eigenvalues. Lower order EOFs represent the large scale structure of the atmospheric variability, while the higher order EOFs stand for the more local features. Therefore, XPL reduces the influence of the small scale features, while MAH increases their significance. Together these measures of similarity account for a considerable range of approaches for identifying analogs.

3.2 Selection process

We consider precipitation at a given target-day (ψ^0) as determined by the state of the atmosphere on that day and the preceding 7 days. Such a sequence is called a target-sequence $\vec{\psi} = (\psi^7, \dots, \psi^0)$. The search for the closest analog is conducted in the 61-day interval centered at the target-date, for all years except for the target-year. For every target-sequence $\vec{\psi}$ and any possible sequence $\vec{\varphi} = (\varphi^7, \dots, \varphi^0)$ the following expression is evaluated:

$$S_{\gamma_p, d}(\vec{\psi}, \vec{\varphi}) = \sum_{k=0}^7 \gamma_p(k) d(\psi^k, \varphi^k) \tag{6}$$

where ψ^k and φ^k stand for single-day-patterns of the sequences ($0 \leq k \leq 7$) and d represents a measure of similarity (see Subsect. 3.1). The weighting sequences $\gamma_p(k)$ are lag-weights that depend on the index $p = 1, \dots, 9$, shown in Fig. 2. They control to what extent preceding days influence the analog search and thereby the use of different sequences will result in different analogs. This variety in addressing bygone days is meaningful as

different quantities (e.g., totals of heavy precipitation events or the length of dry spells) are expected to feature a different dependence on the evolution of the atmosphere. Low-indexed sequences give a higher weighting for the days shortly preceding the target day than for other days in the analysis. High indexed sequences also attach most weight to days close to the target day, but comparable large weights are applied to days constituting a longer time period. The pattern-sequence minimizing the left hand side of Eq. (6) is selected as the analog to the target-sequence. The precipitation field corresponding to its eighth day (φ^0) is taken as the downscaled precipitation field of the target-day (ψ^0).

4 Results

Our main focus is on weather features such as the occurrence of wet and dry spells over a period of some decades (i.e., quantities that should be properly reproduced by a downscaling technique). The modeling of the interannual maximal lengths of dry spells (Subsect. 4.2.1), or the assessment of AM’s skill to estimate precipitation events by comparison to the persistence method (Subsect. 4.2.4), characterize desirable features of a model that links large scale processes to local scale precipitation. However, high model performance regarding these quantities is not a prerequisite when considering the application as a downscaling-technique to generate climates on regional scales.

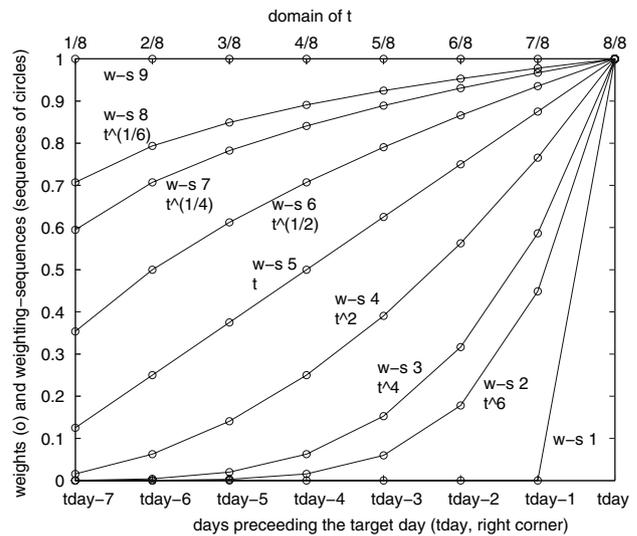


Fig. 2 Weighting-sequences used to identify the analog to the target-sequence. The rightmost mark on the abscissa refers to the target day, those to its left to preceding days. Associated y-values are applied as weights to the state-vectors of these days

4.1 Monthly precipitation

Figure 3 shows AM's skill at replicating regionally averaged, monthly precipitation calculated from daily totals. Panels are based on different settings listed in the titles ('setting' stands for a combination of similarity measure, weighting-sequence and EOF-truncation). Using the setting of the left panel, observed anomalies in CA are well reproduced, with correlation of about 0.9. Correlation is significant for all stations at the 5% significance level. The station averaged long term mean is underestimated by 13% and its variability by 9%. The performance is poorer for XPL for which correlation drops to 0.7. The long term mean and the interannual variance are underestimated again.

The second row in Fig. 3 refers to EA with the same settings as in the upper left panel. Correlations are significant for 93% of the stations and the correlation for regionally averaged precipitation is about 0.7. Long term mean and interannual variance are underestimated by 8% and overestimated by 10%, respectively. The performance in EA is poor when XPL is used. Hence, regardless of the region, the setting in the left hand panels yields a better reproduction of the observations.

Taking both regions together, there are 720 different settings composed of similarity measures, weighting-sequences and EOF-truncation. The left panel of Fig. 4 shows the correlation across all settings for CA. Except for XPL, all similarity measures show the poorest correlation

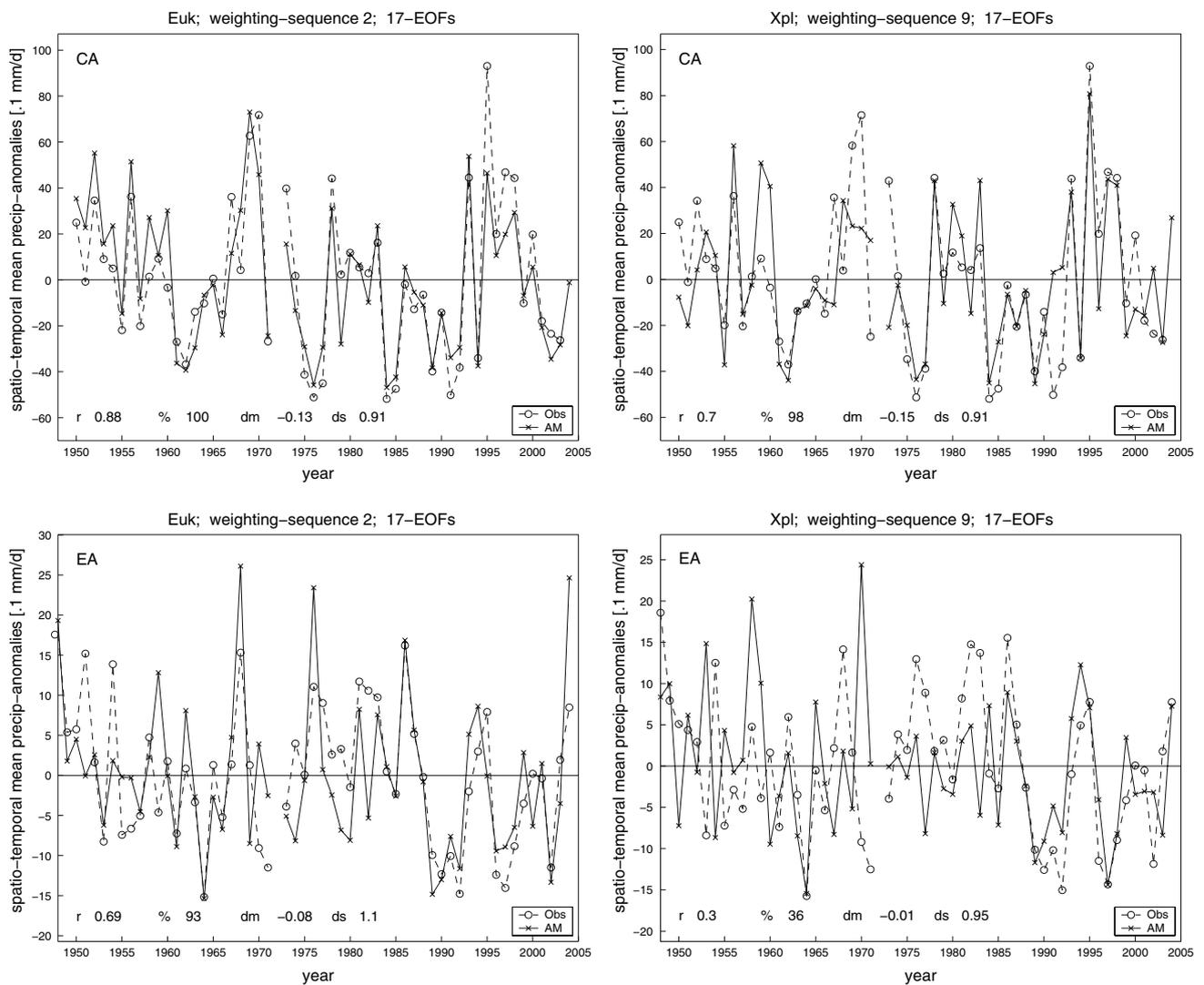


Fig. 3 Observed and downscaled January precipitation for CA (upper panels) and EA (lower panels). Settings are indicated in the panel-titles. Numbers close to the panel-bottoms refer to: correlation coefficient; percentage of stations exhibiting significant correlation

values ($\alpha = 0.05$); difference in mean between simulation and observation, relative to the observed value and the ratio of the simulated to the observed variance

for low EOF-truncation and high indexed weighting-sequences. High values of the Pearson's correlation coefficient are found for EOF-truncation at medium to high dimensions and for low to medium indexed weighting-sequences. Hence, AM performs better in simulating monthly precipitation when small scale atmospheric features are included and when no, or only little information about the past evolution is considered. In case of XPL almost no dependence on the EOF-truncation is found. This behavior is caused by assigning smaller weights to small scale details, that is, higher order EOFs contribute very little to the similarity measure. Figure 4 further indicates that performance saturates for EOF-truncations larger than about 17 (i.e., approximately 85% of the total variance in both regions). EUK, XPL and SUM generally underestimate the amount of total precipitation and variability (not shown). COS and MAH overestimate both the mean amount and variability. Based on weighting-sequence 4, Cos is closest to the observed total but overestimates variability by about 20%. MAH reaches for low indexed weighting-sequences twice the observed amount for the total and temporal variability (not shown).

Results for EA (Fig. 4, right panel) are similar to CA but less distinct and correlation is poorer. Regarding correlation, MAH performs closer to EUK, SUM and COS than to XPL, which shows poorest results. MAH still overestimates both total precipitation and variability, albeit far less pronounced than for CA (not shown).

Overall, EUK, SUM and COS results in simulated monthly precipitation closer to the observation, whereas XPL and MAH perform worse. Thus, enhancing local details as in MAH or suppressing them as in XPL, is not beneficial. The

performance of Cos falls behind those of EUK and SUM but is better than that of XPL and MAH, suggesting that the angle between state-vectors already carries enough information to achieve meaningful results. Performance for EA is clearly reduced compared to CA, which will be discussed later.

4.2 Daily precipitation based quantities

In the following we evaluate AM's ability to reproduce the year to year observed maximum length of dry spells (Subsect. 4.2.1) and compare the fraction of correct categorical precipitation estimates based on AM to that of a random approach (Subsect. 4.2.2). We further discuss persistence by means of climatological frequencies of dry and wet spells (Subsect. 4.2.3), followed by a brief testing of AM's 'forecasting' (quotes are discussed below) power.

4.2.1 Maximum dry spell lengths

Figure 5 shows results which are based on the same settings as in Fig. 3. In CA, EUK achieves a correlation of 0.64 and results are significant at single station series (84%). These values drop to 0.60 and 61% in EA, but results should still be regarded as considerable. For CA XPL, reaches 0.46 for the correlation and simulations are significant at 45% of all stations and overall there is not much difference in performance regarding the two study regions. Correlation all over the settings as shown in Fig. 4 are generally in agreement with the findings based on monthly totals (not shown).

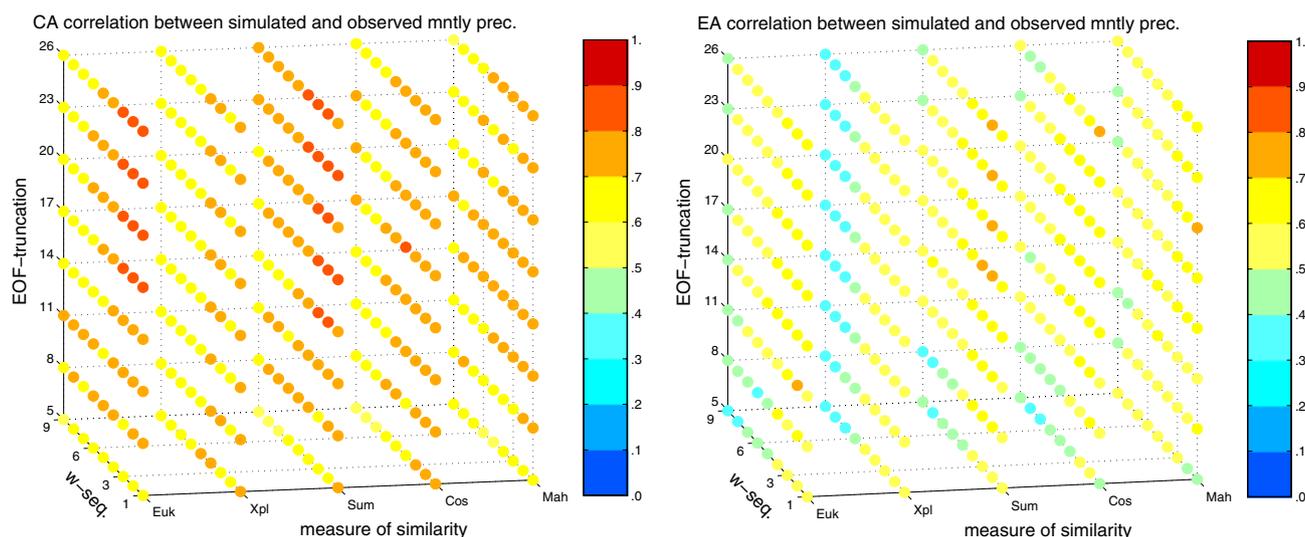


Fig. 4 Correlation between simulated and observed monthly, station averaged precipitation series. x-axis: distance measures; y-axis: weighting-sequences (see Fig. 2); z-axis: EOF-truncation; left/right panel: CA/EA

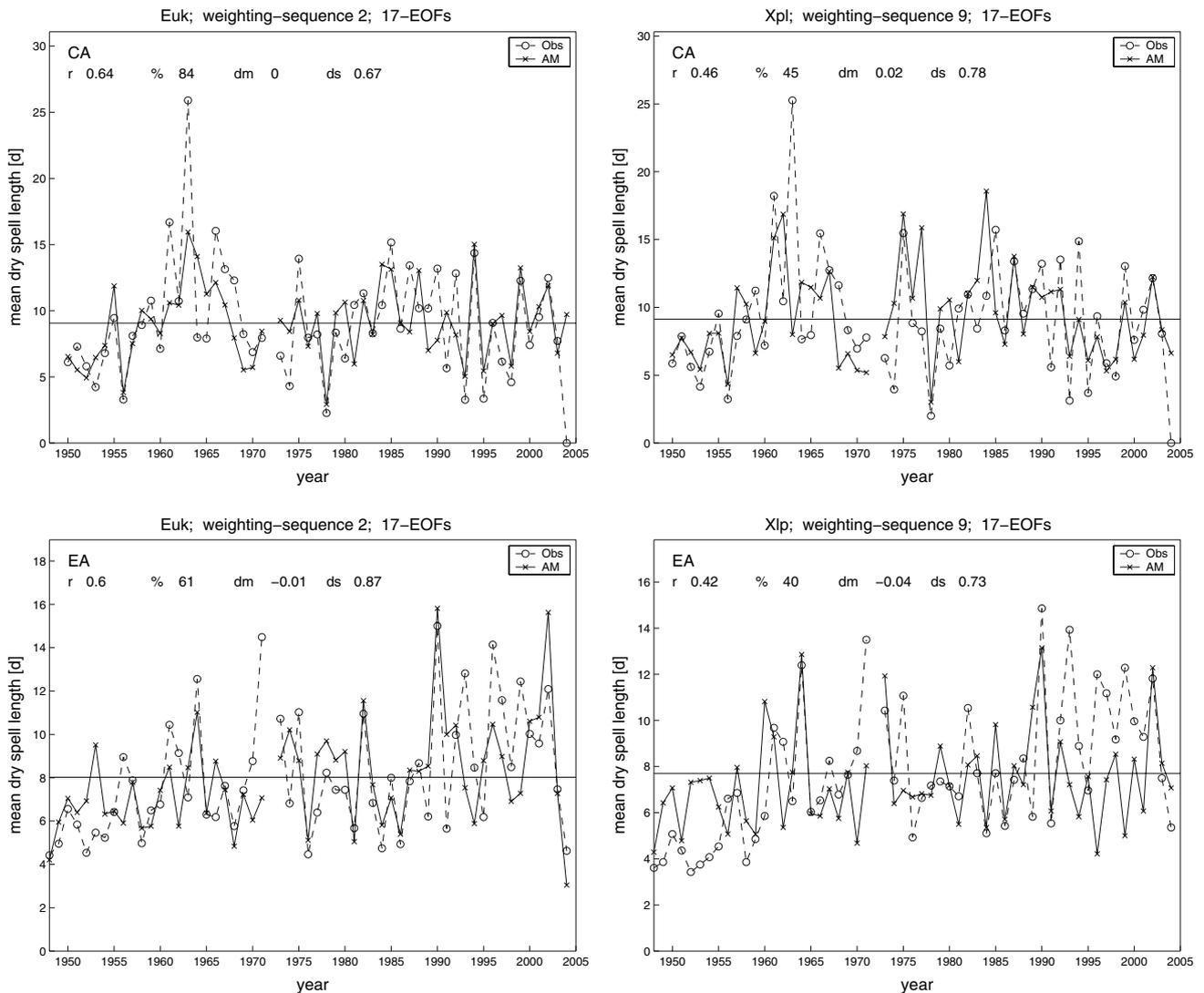


Fig. 5 Observed and downscaled maximal length of dry spells in January for CA (upper panels) and EA (lower panels). Numbers close to the panel-tops refer to: correlation coefficient; percentage of

stations exhibiting significant correlation values ($\alpha = 0.05$); difference in mean between simulation and observation, relative to the observed value and the ratio of the simulated to the observed variance

4.2.2 Fraction of correct categorical precipitation estimates

Even though an accurate simulation of daily precipitation amounts from the large scale atmosphere is not an ultimate goal of downscaling, when being used to produce regional climates, it is a desirable feature. The probability of estimating correctly the precipitation amount is of high practical importance. In the following, daily precipitation is classified into six categories: ‘zero’ (no precipitation), ‘light’ (below the 25th percentiles), ‘medium’ (between the 25th and 75th percentiles), ‘enhanced’ (75th to 90th), ‘strong’ (90th to 95th) and ‘heavy’ (above the 95th percentiles). Categories as simulated by AM are validated

against the observations. Skill is compared to that in the random selection approach.

Figure 6 shows the fraction of a correct categorical classification using AM divided by the fraction due to chance as obtained in the random approach, which is the fraction of correct choices to all possible choices. The panels refer to ‘light’, ‘enhanced’ and ‘heavy’ categories and are based on Euk. In CA (EA) ‘light’ precipitation events are correctly classified from 10 to 14% (10–12%). Based on the random approach 9% of the ‘light’ events are correctly classified. Depending on the applied lag-sequence and EOF truncation, ‘enhanced’ events (middle panels) are correctly identified by AM from 14 to 21% (8–14%) for CA (EA) compared to about 5% by the random approach.

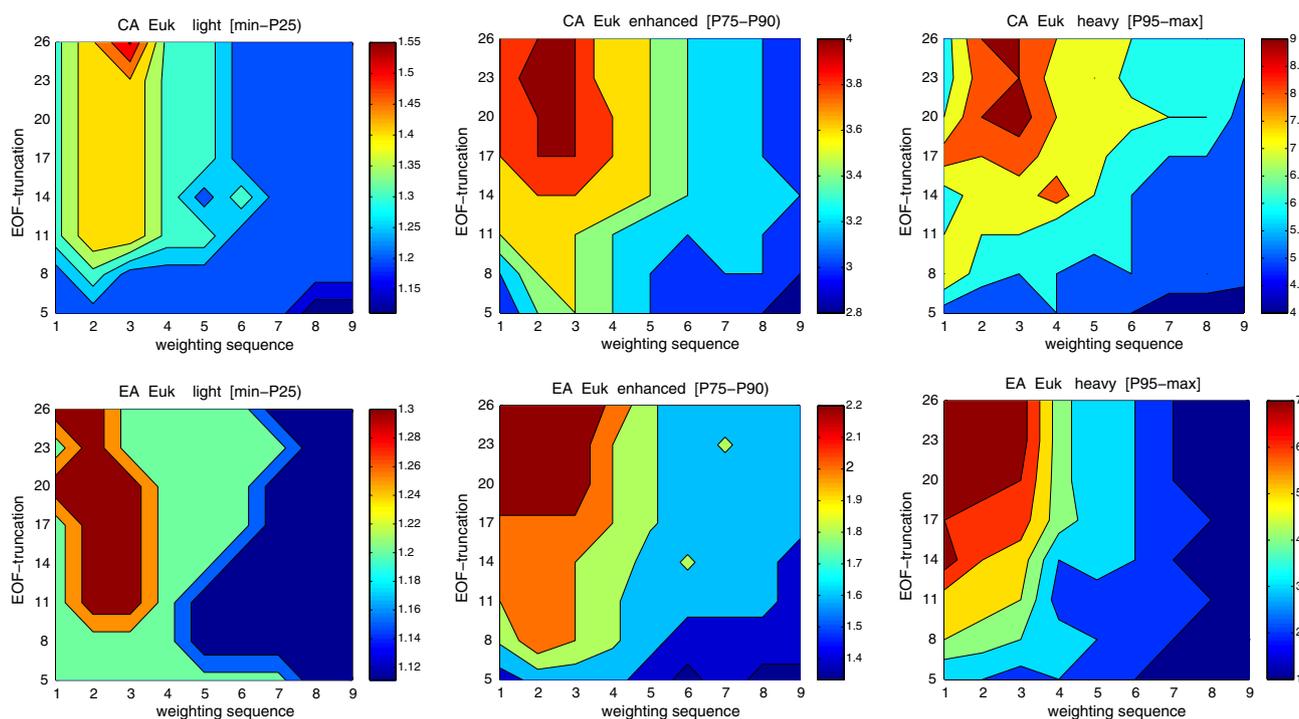


Fig. 6 Probability of a correct categorical precipitation estimate for AM based on EUK, divided by that of the random approach (see text). *first/second row: CA/EA. left to right: precipitation classes: [min,*

P₂₅), [P₇₅, P₉₀] and [P₉₅, max]. x-axis: weighting-sequence; y-axis: EOF-truncation; note the different color bar ranges

Hence, in CA the AM skill is about four times that of the random approach. Ratios become larger for the ‘heavy’ case (right panels).

‘Skill-patterns’ as shown in Fig. 6 are computed for all measures of similarity (not shown). AM generally outperforms the random approach for all classes. For all similarity measures, correct categorical estimates by AM are more probable when based on low to medium indexed weighting-sequences and medium to high values of EOF-truncation. The ratio of successful estimates by AM and by the random approach increases towards ‘heavy’ events, as indicated by the findings for EUK. The shapes of the skill-patterns for EUK, SUM and Cos share similarities and the achieved values are higher than those for XLP or MAH. In both regions XPL shows a pronounced dependency on the lag-sequences but almost no dependency on EOF truncation. Again, AM performs better in CA than in EA, which will be discussed later.

4.2.3 Distribution of dry and wet spells

Frequency distributions of dry and wet spells summarize persistent features of weather variability. Ideally these features should be correctly reproduced by a downscaling technique. In the following we examine AM skill in simulating frequency distributions of 3, 6 and 9 consecutive dry days and of 2, 3 and 4 consecutive wet days. Averaged

over all stations in CA, dry spells of length 3, 6, and 9 days occur 250, 90, and 45 times, respectively, in the study period. The corresponding numbers in EA are 240, 105 and 55, indicating that EA is drier in January than CA. This is reflected by the occurrence of wet spells as well. The number of 2, 3, and 4 day wet spells is approximately 200, 100 and 60, respectively, in CA and 50, 15, and 6 in EA. Winter is the driest season in EA and the wettest season in CA (e.g., Auer et al. 2001; Pandey et al. 1999). Figure 7 shows AM’s ability to model dry and wet spell frequencies for CA, based on weighting-sequence 5. Generally, frequencies of dry spells are underestimated. This negative bias increases with the length of spells. However, the simulations of dry spell frequencies using EUK, SUM and to a lesser degree Cos, appears reasonable. The performance improves when more information from the past evolution is included in the analysis. This improvement is common to all similarity measures and is particularly pronounced for 6 and 9 day-spells (not shown). EOF truncation around 17 allows for best results. MAH and XPL show pronounced negative biases reaching about 40% (not shown). This ought to be related to how the frequency distributions of precipitation are reproduced. EUK, XPL and SUM tend to slightly underestimate percentiles of observed precipitation, Cos overestimates somewhat and MAH strongly overestimates medium to high percentiles. This causes too many cases of rain and too

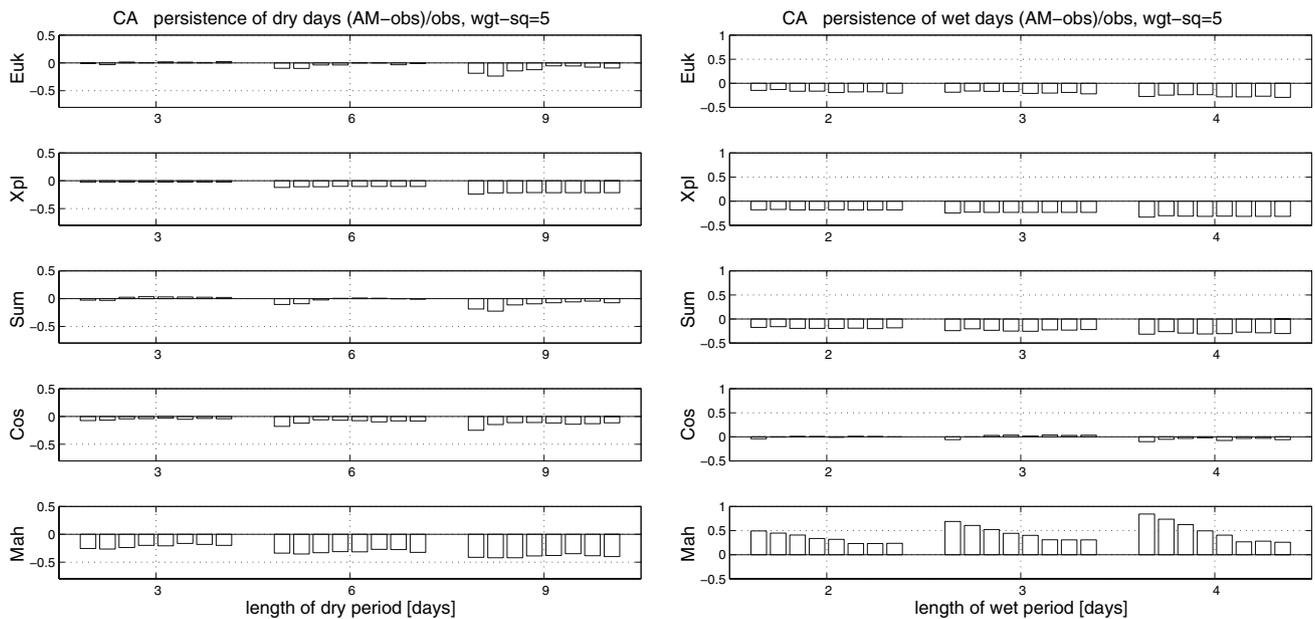


Fig. 7 Assessment of AM's ability to reproduce dry (*left*) and wet (*right*) spell frequencies based on weighting-sequence 5. The subpanels show the measures of similarity and display ($f_{AM}-f_{obs}/f_{obs}$ with f , the frequency of occurrence). Groups of bars (*left to right*) refer

to 3-, 6-, 9-day dry spells and 2-, 3-, and 4-day wet spells respectively. Bars within these groups (*left to right*) correspond to EOF-truncations from 5 to 26 in steps of 3

many strong to heavy events, which in turn suggests less dry spells. This is a serious caveat and hence MAH should not be used in the analysis of precipitation unless there is a particular reason to do so.

The right panel in Fig. 7 shows the results for wet spells based on weighting-sequence 5 again. Frequencies of wet periods are somewhat underestimated by EUK, XPL and SUM, closely matched by Cos and strongly overestimated by MAH. There is an increase in performance up to weighting-sequence 3, followed by a decrease, most apparent for EUK and SUM (not shown). This seems to reflect the shorter lengths of the considered wet spells compared to the dry spells, making more distant days less important. XPL shows not much dependence on weighing sequences and overall Cos performs best. EOF-truncation around 17 is a reasonable choice again. This remains essentially the same for EA, albeit at poorer performance. One difference to CA, worthwhile mentioning is that frequencies of wet periods are best reproduced by consideration of medium to high indexed weighting sequences.

The rather reasonable performance of Cos in modeling wet spell frequencies in both regions, indicates that the shape of the large scale atmospheric state is particularly important. Cos addresses high similarity to states that are represented by parallel state vectors, regardless of their difference in length, which is different from EUK and SUM, emphasizing the shape of atmospheric patterns over their strength when simulating wet spell length.

4.2.4 Comparison to persistence method

AM was also used as a forecasting tool (see e.g., Soucy 1991 for the Canadian Metrological Center or Livezey et al. 1994 for the United States). Obviously, it is not intended to recommend AM over today's numerical weather forecasting approach based on ensembles of model-runs. However, the ability to predict precipitation is an important feature of today's forecasting models and hence, AM can still serve its purpose as a bench mark that has to be surpassed. As such, it is of interest to compare AM's 'forecasting'-skill to the persistence method (PM). PM predicts tomorrow's weather from today's weather by assuming conditions will not change. PM works well for regions where weather patterns do not change quickly.

In the following we compare AM to PM by evaluation of the associated daily based correlation coefficients between estimation and observation. We have put forecasting in quotes above as AM is provided with the best large scale field-estimates available, the reanalysis data. This shall be understood and hence, findings represent an upper limit of performance.

Figure 8 compares AM to PM at all stations for a specific setting (see title). AM almost always outperforms PM at quite high correlation values. As the stations are numbered eastward (see Fig. 1), those in the left of the panel tend to be closer to the Pacific ocean at lower altitudes, while the ones in the right are more likely to be located

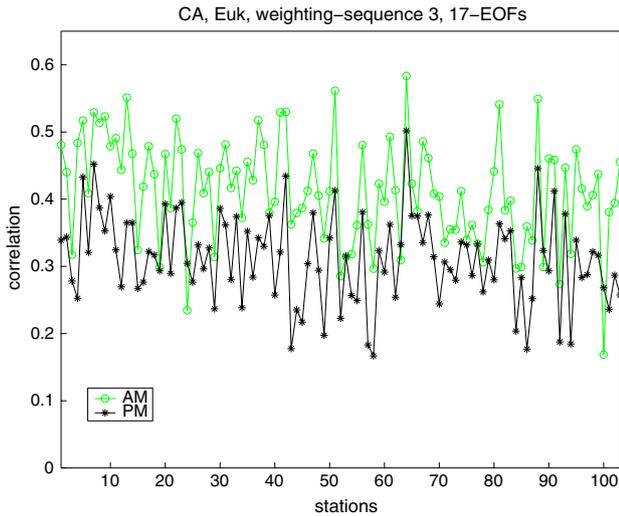
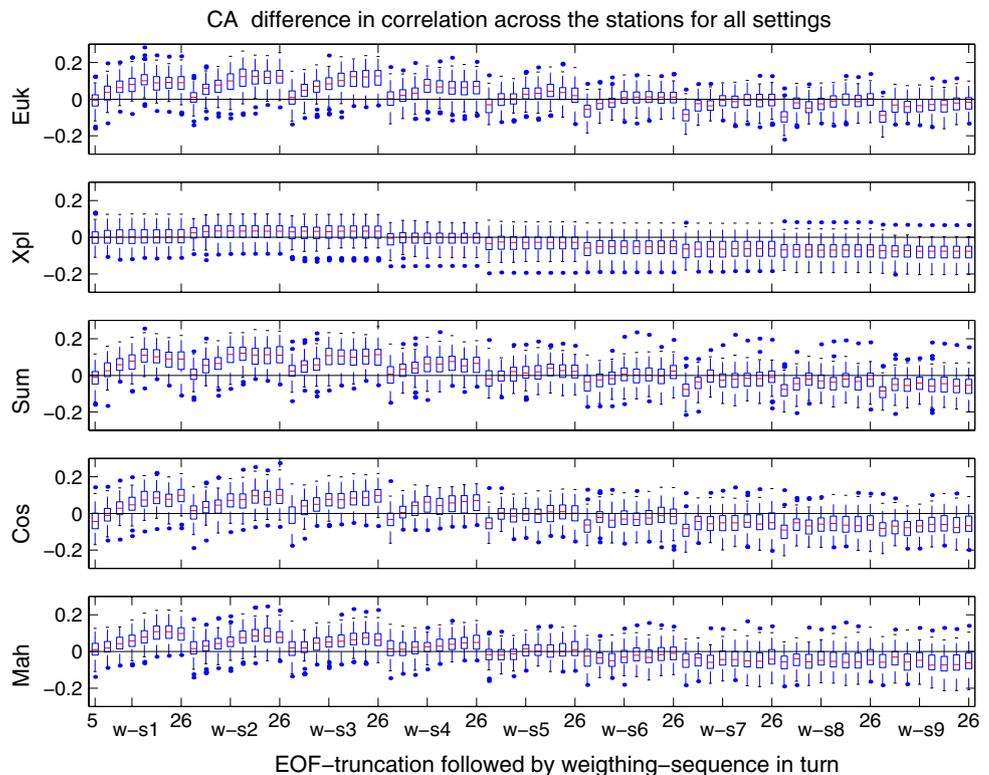


Fig. 8 Correlation coefficients across the stations in CA as modeled by AM (setting in the title) and PM

within the Sierra Nevada at higher elevations. AM shows no prominent dependency on elevation. An analysis (not shown) of the reproduction of precipitation distributions across three altitudinal ranges (indicated in Fig. 1) has not identified a dependence either. Hence, the chosen fields (SLP and SH7) demonstrate similar levels of utility in modeling local scale precipitation at stations of different elevation and thus across different climatic zones (as indicated in Fig. 1). This is different from EA, where

Fig. 9 The figure displays the distribution of $r_{AM}-r_{PM}$ across all stations in CA as boxplots, where r is the correlation coefficient. The five subpanels refer to the similarity measures. The x -axis strings all settings from weighing-sequence 1 ($w-s1$) and EOF-truncation at 5 weighing-sequence 9 ($w-s9$) and EOF-truncation at 26



stations in the north-eastern lowlands exhibit particular weak correlations (not shown), a region known for its particular precipitation climate (Auer et al. 2001).

Figure 9 provides a comprehensive survey of the difference between AM and PM based correlation coefficients across all possible settings and all stations located in CA. Basically, up to weighing-sequence 4, AM reaches values of correlation higher than PM. From weighing-sequence 6 onward this picture reverses. Except for XPL, all measures of similarity depend on EOF-truncation. Performance increases up to 17 EOFs, remains the same or slightly decreases at higher EOF-truncation. Weighting-sequence 2 is a reasonable choice for all measures. Hence, estimates of the target-day’s precipitation pattern across the stations benefit most from the use of the target-day and the preceding two to three days in terms of weighing-sequence 2. Inclusion of longer elapsed states degrades skill. Among the similarity measures XPL performs poorest. For EA (not shown) these features remain generally in effect, but there is a pronounced overall drop in performance, that let PM outperform AM at almost all settings.

5 Discussion

The study provides a comprehensive assessment of AM in two regions that are characterized by rather distinct precipitation climates. The assessment is done across a large

variety of settings covering reasonable measures of similarity and broadly based approaches to regard the predictor-fields and their temporal evolution. For every single setting a temporal cross-validation was performed, resulting in about 40 k simulated months. Hence, derived results should be significant and settings, identified to allow AM for considerable performance, are expected to be a rather reasonable starting point for other regions as well.

Most obviously and for all considered tasks, downscaled precipitation is at a higher level of accordance to observations in CA than in EA. That should be attributed to the closeness of the link between the large scale state of the atmosphere and the local scale precipitation. California is located at the US west coast next to the Pacific Ocean and the Sierra Nevada is perpendicular oriented to the westerlies. Proximity to the ocean, storms approaching frontally Sierra Nevada's mountain sides and orographic lifting, exert strong influence on local scale precipitation. The study region in the European Alps, on the other hand, is about 1000 km away from the Atlantic Ocean and parallels the westerlies. Large distance and a considerable variety of air masses approaches to the region perturb the link between the scales, causing AM to achieve higher performance in CA. However, most of AM's characteristics remain in effect for both regions, putting statistical weight on the findings. First, throughout all tasks, best performance was achieved at EOF-truncation accounting for 85–90% of explained variability. Inclusion of more EOFs does not enhance results. This could be related to the decreasing statistical significance of higher order EOFs (von Storch and Hanneschöck 1985).

Recommendation regarding weighting-sequence i.e., the extent to which inclusion of more distant past atmospheric states is beneficial, depends on the considered task. Mapping the run of monthly totals or issuing categorical estimates of the intensities of precipitation events is generally most successful when based on weighting-sequence 2 and 3 (sixth and fourth power weighting, see Fig. 2). Quantities, dealing with persistence are found to be better reproduced when addressing more weight to the more distant past, depending on the considered endurance.

Measures of similarity that introduce different weights to the components fall behind in performance. XPL underweights local detail of the atmospheric state and hence, cannot take advantage of higher order EOFs. MAH reverses this effect and strongly overestimates medium to high daily percentiles, yielding too large monthly sums, too few dry spells and too many wet spells. Hence, MAH should not be considered for downscaling and XPL cannot be recommended either, as its performance falls behind the others. MAH and XPL are two special cases to introduce different weights to the EOF coordinates as formulated in generality by Zorita and von Storch (1999). The attachment of

different weights to the components is a transformation that may significantly alter the EOF assembly. Findings show that neither a monotonically increasing (MAH) nor a decreasing (XPL) sequence of weights is of help. So, if a component-sensitive weighting is desired, considerable effort is required to solve properly formulated optimization procedures that define these weights. Cos overestimates medium to high percentiles and performs at some tasks weaker than EUK and SUM, but has advantages over them for wet-spell length (particularly together with weighting-sequence 5) and when modeling the maximal length of dry spells. Cos identifies the most parallel state vectors as analogs (see Subsect. 3.1). So, the mentioned features (e.g., wet spell length) are shared more by sequences of days having almost parallel state vectors. Such states are similar in terms of pattern shapes but can have large differences in length and hence would not be selected by EUK or SUM. However, pronounced difference in length should impact the totals. EUK and SUM perform very comparable. So, both approaches quantifying similarity between atmospheric states lead to rather reasonable results.

This study is an extension of previous work (Zorita et al. 1995; Zorita and von Storch 1999) and was motivated by problems raised therein. It addresses these issues by using a broad setup. Findings of this study may serve as guidelines when using AM for downscaling of daily precipitation. It also provides the opportunity to demonstrate the achievements of more complex models by comparison to AM, based on recommended settings.

Acknowledgments This work was supported by a Visiting Fellowship to Canadian Government Laboratories awarded to C. Matulla through NSERC. J. Wang was supported by CFCAS. We are thankful to H. Kuhn who has enabled us to a trouble free processing of our calculations, to E. Watson, V. Kharin and D. Bray for stimulating suggestions and comments on this study. Furthermore we are grateful to two anonymous reviewers who helped to strengthen the manuscript.

References

- Auer I, Böhm R, Schöner W (2001) Austrian long-term climate 1767–2000—multiple instrumental climate time series from Central Europe. Technical Report 25, Central Institute for Meteorology and Geodynamics, Vienna, pp 147
- Baur F (1951) Extended-range forecasting by weather types. *Compendium of meteorology*. Am Meteor Soc, pp 814–833
- Elliot RD (1951) Extended-range weather forecasting. *Compendium of meteorology*. Am Meteor Soc, pp 834–840
- Gleason BE (2002) National climate data center data documentation for data set 9101, global daily climatology network v1.0. Technical report, National Climatic Data Center, Asheville
- Gruza GV, Ran'kova EY (1986) Monitoring and probability forecast of monthly and seasonal variations in hemisphere atmospheric processes. In: *Proceedings of First WMO workshop on the diagnosis and prediction of monthly and seasonal atmospheric variations over the globe*, vol II of long-range forecasting res.

- Rep. Ser. 6, WMO/TD 87. World Meteorological Organization, Geneva, pp 682–688
- Hamill TM, Whitaker JS (2006) Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Mon Wea Rev* 134:3209–3229
- IPCC (2001) Climate Change 2001: Working Group I: The scientific basis. Appendix 10.4: Examples of downscaling studies. Cambridge University Press, Cambridge
- 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) The NCEP/NCAR reanalysis project. *Bull Am Meteor Soc* 77:437–471
- Livezey RE, Barnston AG (1988) An operational multifield analog/antianalog prediction system for United States seasonal temperatures. Part 1: System design and winter experiments. *J Geophys Res* 93:10953–10974
- Livezey RE, Barnston GV, Anthony G. and Gruza, Esther Ya. Ran'kova (1994) Comparative skill of two analog seasonal temperature prediction systems: objective selection of predictors. *J Clim* 7:608–615
- Lorenz EN (1969) Atmospheric predictability as revealed by naturally occurring analogs. *J Atmos Sci* 26:636–646
- Pandey GR, Cayan DR, Georgakakos KP (1999) Precipitation structure in the Sierra Nevada of California during winter. *J Geophys Res* 104:12019–12030
- Ropelewski CF, Halpert MS (1986) North American precipitation and temperature patterns associated with the El Niño/Southern oscillation (ENSO). *Mon Wea Rev* 114(12):2352–2362
- Schöner W, Auer I, Böhm R, Thaler S (2003) Quality control and statistical characteristics of selected climate parameters on the basis of daily values in the face of extreme value analysis (German). In: Kromp-Kolb H, Schwarzl I (eds) *StartClim–Start* Project: First analysis of extreme weather events and their impacts on Austria, vol 1. Institute of Meteorology and Physics, BOKU - University of Natural Resources and Applied Life sciences, Vienna, p 52
- Schonher T, Nicholson SE (1989) The relationship between California rainfall and ENSO events. *J Clim* 2:1258–1269
- Soucy D (1991) Revised users guide to days 3–4–5 automated forecast composition program. Canadian Meteorological Centre, Environment Canada
- Toth Z (1991) Intercomparison of circulation similarity measures. *Mon Wea Rev* 119:55–64
- van den Dool H (1994) Searching for analogs, how long must we wait? *Tellus* 46A:314–324
- von Storch H, Hanneschöck G (1985) Statistical aspects of estimated principal vectors (EOFs) based on small sample sizes. *J Clim Appl Meteor* 24:716–724
- von Storch H, Zwiers FW (1999) *Statistical analysis in climate research*. Cambridge University Press, Cambridge, 528 pp
- von Storch H, Zorita E, Cubasch U (1993) Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. *J Clim* 6:1161–1171
- Yamnor WS, Draper BA, Beveridge JR (2002) Empirical evaluation methods in computer vision, volume 50 of series in machine perception and artificial intelligence. In: Christensen H, Phillips J (eds) *Chapter analyzing PCA-based face recognition algorithms: eigenvector selection and distance measures*. World Scientific Press, Singapore
- Zorita E, Hughes J, Lettenmaier D, von Storch H (1995) Stochastic characterisation of regional circulation patterns for climate model diagnosis and estimation of local precipitation. *J Clim* 8:1023–1042
- Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *J Clim* 12:2474–2489