

A novel approach to statistical downscaling considering nonstationarities: application to daily precipitation in the Mediterranean area

E. Hertig,¹ and J. Jacobeit¹

Received 24 July 2012; revised 10 December 2012; accepted 12 December 2012; published 30 January 2013.

[1] In the present study, nonstationarities in predictor–predictand relationships within the framework of statistical downscaling are investigated. In this context, a novel validation approach is introduced in which nonstationarities are explicitly taken into account. The method is based on results from running calibration periods. The (non)overlaps of the bootstrap confidence interval of the mean model performance (derived by averaging the performances of all calibration/verification periods) and the bootstrap confidence intervals of the individual model errors are used to identify (non)stationary model performance. The specified procedure is demonstrated for mean daily precipitation in the Mediterranean area using the bias to assess model skill. A combined circulation-based and transfer function–based approach is employed as a downscaling technique. In this context, large-scale seasonal atmospheric regimes, synoptic-scale daily circulation patterns, and their within-type characteristics, are related to daily station-based precipitation. Results show that nonstationarities are due to varying predictors–precipitation relationships of specific circulation configurations. In this regard, frequency changes of circulation patterns can damp or increase the effects of nonstationary relationships. Within the scope of assessing future precipitation changes under increased greenhouse warming conditions, the identification and analysis of nonstationarities in the predictors–precipitation relationships leads to a substantiated selection of specific statistical downscaling models for the future assessments. Using RCP4.5 scenario assumptions, strong increases of daily precipitation become apparent over large parts of the western and northern Mediterranean regions in winter. In spring, summer, and autumn, decreases of precipitation until the end of the 21st century clearly dominate over the entire Mediterranean area.

Citation: Hertig, E., and J. Jacobeit (2013), A novel approach to statistical downscaling considering nonstationarities: application to daily precipitation in the Mediterranean area, *J. Geophys. Res. Atmos.*, 118, 520–533, doi:10.1002/jgrd.50112.

1. Introduction

[2] Within the scope of projecting future climate change, general circulation models (GCMs) are commonly used to assess changes resulting from further increases of atmospheric greenhouse gases. Different types of regionalization techniques have been developed to infer regional to local information below the skillful scale of the GCMs. Statistical downscaling provides a computationally inexpensive technique that can be adapted for a wide range of applications. Statistical downscaling approaches are based on establishing statistical relationships that link a set of large-scale atmospheric variables (predictors) to regional or local climate variables (predictands) over an observational period. For an overview on statistical downscaling techniques, see studies by *Maraun et al.* [2010] and by *Wilby and Wigley* [1997]. In this regard, validation of the statistical

models is of utmost importance: the established statistical relationships are validated during a period independent from the calibration period and are subsequently used to project the future response of regional climate to climate model changes of the large-scale variables.

[3] As validation procedures split-sampling [appropriate when long observational records are available, e.g., *Busuioc et al.*, 1999], cross-validation techniques [particularly when only shorter records are available, e.g., *Murphy*, 1999], stratified validation [e.g., using subsets with dry years and wet years, respectively, *Wilks*, 1999], and statistical model ensembles [using multiple calibration/ verification periods to account for nonstationarities in the predictor–predictand relationships, *Hertig and Jacobeit*, 2008] have been applied. Beyond these general techniques to evaluate the statistical model skill using independent data, the downscaling relationships should be analyzed for their physically reasonable linkages, as previously undertaken in the pioneering work of *von Storch et al.* [1993]. Furthermore, comparisons between different statistical downscaling approaches [e.g., *Zorita and von Storch*, 1999], between statistical and dynamical downscaling [e.g., *Murphy*, 1999], and the use of pseudorealities [*Vrac et al.*, 2007] can be adopted.

¹Institute of Geography, Universitaetsstrasse 1a, 86159 Augsburg, Germany.

Corresponding author: E. Hertig, Institute of Geography, Universitaetsstrasse 1a, 86159 Augsburg, Germany. (elke.hertig@geo.uni-augsburg.de)

©2013. American Geophysical Union. All Rights Reserved.
2169-897X/13/10.1002/jgrd.50112

[4] Despite extensive efforts to measure and evaluate the performance of statistical downscaling models, little focus has been given to the handling of nonstationarities in the predictors–predictand relationships. According to *Wilby* [1998], nonstationarities in statistical downscaling models may be traced back to three underlying factors: an incomplete set of predictor variables, inadequate calibration periods, or situations in which the climate system structure(s) changes through time. The issue of an incomplete set of predictor variables has been reduced in recent years due to advancements in GCM's ability to model predictor variables more accurately. However, there are certainly still major improvements necessary regarding this matter. Inadequate calibration periods may be improved using longer data sets, whereas the aspect related to nonstationarities inherent to the climate system perhaps forms the major obstacle.

[5] Nonstationarities are inherent characteristics of the climate system and can be observed on different temporal and spatial scales: one of the best known atmosphere–ocean interactions is the El Niño–Southern Oscillation, which influences the climate of many regions around the world. In this context, a major climatic shift occurred in the 1970s in the Pacific region [*Trenberth*, 1990; *Miller et al.*, 1994; *Wang*, 1995], which consequently led to an alteration of the El Niño–Southern Oscillation teleconnections [to the extra-tropics, see e.g., *Greatbatch et al.*, 2004; to the Mediterranean area, e.g., *Mariotti et al.*, 2002]. Beyond these nonstationarities operating on decadal to multidecadal timescales, the variability of atmospheric regimes is an important source for nonstationarities, taking place on interannual to decadal timescales. Thus, the North Atlantic Oscillation (NAO) plays an important role for climate in Europe and the Mediterranean, particularly in winter [*Hurrell et al.*, 2003]. In addition to the NAO, other low-frequency atmospheric circulation patterns have been identified [see e.g., *Barnston and Livezey*, 1987]. All these patterns are subject to temporal variations concerning their location, strength, and mode, and consequently exhibit a varying influence on regional climate. For instance, the NAO was predominantly in its positive mode between the 1970s and early 1990s, causing precipitation decreases in many parts of the Mediterranean area [*Quadrelli et al.*, 2001; *Ben-Gai et al.*, 2001; *Jacobeit et al.*, 2007]. Also, on synoptical scales, nonstationarities are evident because specific circulation patterns exhibit pattern-specific variations. These variations apply to the frequency and persistence of a pattern as well as to within-type variations of pattern-specific temperature and humidity characteristics. Thus, synoptic-scale processes hold another source of nonstationarities, which operate on a daily to interannual timescale.

[6] The present study aims to further investigate nonstationarities within the context of statistical downscaling. In this context, a novel downscaling approach is introduced that takes nonstationarities explicitly into account. In an earlier approach [*Hertig and Jacobeit*, 2008], statistical model ensembles were developed to capture varying predictors–predictand relationships, thereby accounting for nonstationarities. However, no systematic evaluation of the individual statistical ensemble members has been done with regard to specific time periods of changed relationships and the underlying physical processes. This is addressed in the present study by using results from running calibration analysis and bootstrap

confidence intervals. Subsequently, an informed selection of specific downscaling models representing different predictors–predictand relationships becomes possible from this technique. To illustrate the approach, the Mediterranean area is chosen because this area shows a wide range of different climatic characteristics, from humid conditions in the western, northern, and northeastern Mediterranean regions in the wet season (approximately September to May), to arid conditions in the southern and eastern Mediterranean regions in the summer.

[7] The remainder of the article is organized as follows: section 2 describes the data and its preparation for this study; daily station-based precipitation and its regionalization, the observational predictors and their classification into large-scale atmospheric regimes, synoptic-scale circulation patterns and their within-type characteristics, as well as the model predictors based on the latest generation of GCMs and scenarios. Section 3 comments on the statistical downscaling approach to model daily precipitation in the Mediterranean area and, in particular, introduces a novel approach to the validation of statistical downscaling results under explicit consideration of nonstationarities. Results are presented in section 4, beginning with a comparison between observation-based predictors and predictors from the GCM runs. Subsequently, results from the validation approach and a discussion on the reasons for nonstationarities are given. This is followed by the assessment of precipitation until the end of the 21st century. Finally, section 5 includes a discussion of the results and draws some essential conclusions from it.

2. Data and Data Preparation

2.1. Daily Precipitation

[8] Daily station data for the Mediterranean area have been collected from the GLOWA Jordan River Project [Global Change and the Hydrological Cycle, *Kunstmann et al.*, 2006], from the EMULATE project [European and North Atlantic daily to MULTidecadal climATE variability, *Moberg et al.*, 2006], and from the European Climate Assessment and Dataset [*Klein Tank et al.*, 2002]. Unfortunately, there are currently very little data available over the southern Mediterranean area, especially in the countries of Northern Africa. Also, only a few stations are available over the northeastern Mediterranean area (particularly Turkey), but for the other regions, rather good station coverage with daily data has been achieved (see Figure 1).

[9] After testing the completeness [*Moberg*, 2006] of the daily precipitation data, 94 stations, which have less than 4 missing days per 3 month season, could be retained for subsequent analyses. Because one of the main foci of the present study is on statistical downscaling under consideration of nonstationarities in the predictors–precipitation relationships, the homogeneity of the precipitation series has to be tested. This is done in order to avoid artificial nonstationarities due to erroneous station data, which can result from changing measurement conditions (e.g., through station relocations). Therefore, homogeneity is assessed using absolute tests following *Wijngaard et al.* [2003], and a relative test introduced by *Alexandersson* [1986], which is modified to take multiple breakpoints into account.

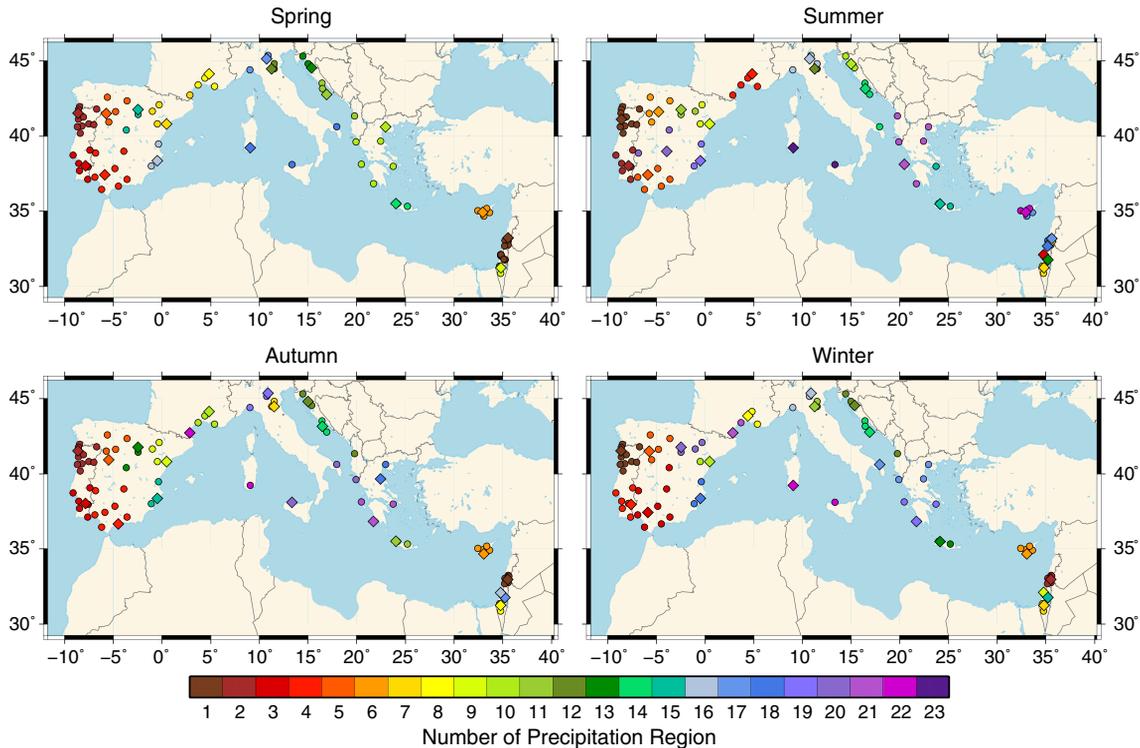


Figure 1. Location of the precipitation stations, their regionalization, and the corresponding representative stations (large symbols) and associated stations (small symbols). For further details, see text.

[10] In order to achieve a regionalization of the precipitation data, principal component analysis [PCA, e.g., Preisendorfer, 1988; von Storch and Zwiers, 1999] is applied to the daily precipitation series in the time period 1961 to 1990 for each natural season (winter, spring, summer, and autumn). The time period 1961 to 1990 is chosen for this analysis because it depicts a period with complete data at all stations. In the present study, S-mode, orthogonally (Varimax criterion) rotated PCAs are carried out. The determination of the number of principal components (PCs) to be extracted follows the approach of Philipp *et al.* [2007] and is based on the criterion that each PC has to be uniquely representative for at least one input variable. Representativeness is assumed when the maximum loading of a variable on a particular PC is at least one standard deviation greater than the other loadings of this variable on the remaining PCs; additionally, this maximum loading has to be statistically significant at the 95% level. Subsequently, the assignment of a precipitation station to a specific precipitation region (i.e., to a specific PC) is done by considering the maximum loading of the station across all the PCs. From the PCA, 18 precipitation regions arise for spring, 23 regions for summer, and 22 regions for autumn and winter. Representative stations for these regions are selected in terms of that station's time series having the absolute highest loading on the corresponding PC. The statistical downscaling models are subsequently derived for the representative stations. The location of all stations, their regionalization by PCA, the corresponding representative stations, and the associated stations are mapped in Figure 1. Note that all subsequent analyses are performed for the whole time period

available for a specific station (with 1950 being the earliest starting year according to the applied reanalysis data).

2.2. Observational and Model Predictors

[11] The predictor variables for the observational time period 1950 to 2010 are obtained from the NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) reanalysis project [Kalnay *et al.*, 1996; Kistler *et al.*, 2001]. As predictors to describe the large-scale atmospheric circulation sea level pressure, and geopotential heights at the 1000, 700, and 500 hPa levels at a $2.5 \times 2.5^\circ$ horizontal resolution are considered. Tests with these predictors show that using 700 hPa geopotential heights works better within the present statistical downscaling procedure in terms of a smaller model bias of statistically downscaled precipitation compared to the other circulation predictors mentioned previously. For that reason, and to account for the different dimensions and timescales on which nonstationarities can occur, geopotential heights of the 700 hPa level in the area 20°N – 70°N , 70°W – 50°E were selected to include large-scale atmospheric regimes that showed interannual to decadal variability. In this context, a large-scale atmospheric regime is defined as an interannually and intrannually recurring state of the atmosphere. The atmospheric regimes are obtained through S-mode PCA of seasonal mean 700 hPa geopotential height fields. Only the first four (autumn and winter) to five (spring and summer) PCs are retained, with overall explained variances of between 62.4% (autumn) and 75.2% (winter). The resulting regimes exhibit high correlations to well-known Northern Hemisphere teleconnection patterns as published by the Climate Prediction Center. This

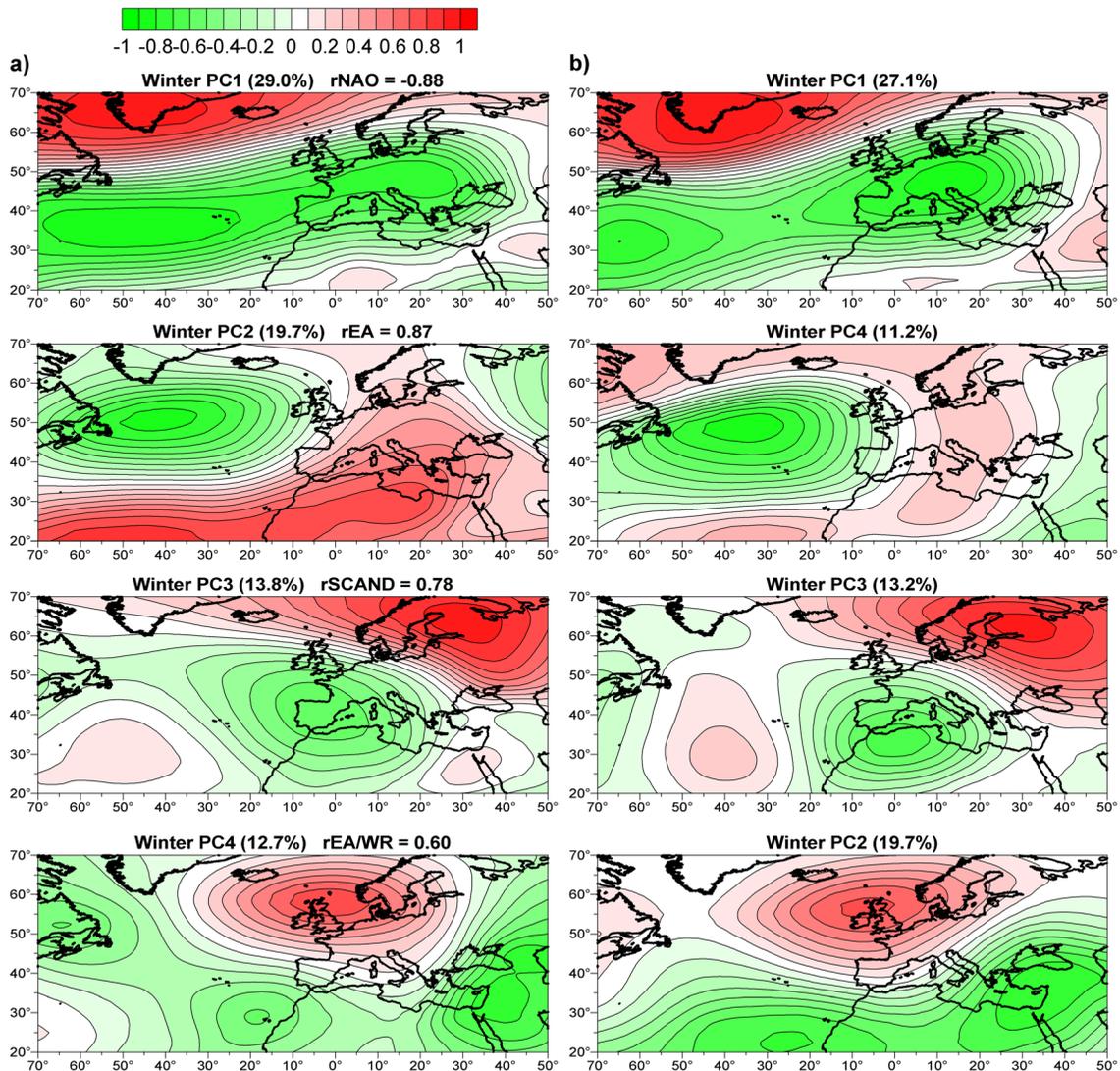


Figure 2. Large-scale atmospheric regimes in winter (December to February), derived from S-mode PCA of seasonal mean geopotential heights of the 700 hPa level in the period 1950 to 2010. Regimes from (a) NCEP/NCAR reanalysis, and from (b) MPI-ESM-LR, run 1.

concerns the NAO, the East Atlantic Pattern (EA), the East Atlantic/Western Russia Pattern (EA/WR), and the Scandinavia Pattern (SCAND). The large-scale atmospheric regimes are further discussed in section 4.1 and are illustrated in Figure 2a for winter.

[12] To account for daily to interannual influences on precipitation, 700 hPa geopotential heights are again used, but in this case, to obtain circulation patterns within station-specific predictor domains. The domain for a specific station in a specific season is selected via correlation analysis of the precipitation time series with all grid points of the large-scale 700 hPa geopotential height fields. After testing different thresholds for the correlation coefficient, correlations that exhibit an absolute coefficient greater than 0.3 are used to cut a rectangular domain incorporating the grid points with these correlations. This empirical threshold is selected because significance levels, even at the 99.9% level, cannot be used as a criterion attributed to the larger number of cases being included in the analysis. Subsequently, T-mode PCA of daily 700 hPa geopotential height fields in the specific domain

are carried out to obtain daily circulation patterns. For the stations considered in the present study, four to seven PCs are extracted (using the same criteria as mentioned previously) with overall explained variances greater than 90% in each case.

[13] Furthermore, 700 hPa relative humidity, zonal and meridional wind components of the 700 hPa level, and convective inhibition (CIN) are used to describe within-type characteristics of the circulation patterns. The selection of these variables is based on previous analyses [e.g., Hertig *et al.*, 2012, Hertig and Jacobelt, 2008] as well as on various studies related to statistical downscaling. The selection of humidity-based predictors in statistical downscaling models includes the question of what kind of variable (absolute or relative humidity) in which atmospheric levels has to be used to get realistic downscaling results. Thus, Charles *et al.* [1999] recommended, for the probability of rainfall occurrence in Australia, the usage of relative moisture (reflecting how close to saturation the atmosphere is) rather than absolute values (reflecting the total water vapor content), taking into account

the increased moisture-holding capacity of the atmosphere under increased temperature conditions. Good results were obtained by *Beckmann and Buishand* [2002] for European stations and by *Hewitson and Crane* [2006] for South African stations using the relative humidity of 700 hPa level as a significant predictor for rainfall occurrence, but the specific humidity of the 700 hPa level as a significant predictor for rainfall amounts. In the present study, tests with models including relative and/or absolute humidity of the 850 and 700 hPa levels, respectively, showed the best model quality when 700 hPa relative humidity is used as the additional predictor to assess changes of daily precipitation amounts. Regarding the wind components, *Cavazos and Hewitson* [2005] show—in a study about the performance of reanalysis variables in statistical downscaling of daily precipitation—that the meridional wind component appears in the list of the top variables, suggesting an influence from surface meridional synoptic systems on precipitation. With regard to convective precipitation, an important requisite is a source of uplift, which can be restated as a requirement of sufficiently small CIN [*Myoung and Nielsen-Gammon*, 2010]. These authors assess that CIN is particularly important for precipitation variability over land areas, and that the existence of a large amount of CIN tends to inhibit the initiation of convection despite substantial conditional instability and moisture availability. Therefore, a proxy for CIN [calculated from reanalysis data; for a detailed description of the index calculation, see *Hertig et al.*, 2012] is also selected as a predictor. Finally, the within-type characteristics are represented by the mean values of the variables over the nine closest grid boxes to the target station. The use of multiple grid boxes around a specific station is adequate to account for possible spatial variations of the main areas of influence, which may occur when changing from reanalysis to GCM conditions.

[14] Global earth system model predictors are taken from a three-member MPI-ESM-LR (Max Planck Institute Earth System Model running on low resolution grid) ensemble with different initial conditions of each run, including historical runs from 1950 to 2005 and RCP4.5 scenario conditions from 2006 to 2100. The representative concentration pathway (RCP) scenario set encompasses emission, concentration, and land-use trajectories. The RCP4.5 scenario leads to an

(compared to the other RCPs) intermediate radiative forcing level of 4.5 W/m^2 by the end of the century [*Van Vuuren et al.*, 2011]. The earth system model MPI-ESM uses the atmospheric ECHAM6 model [*Roeckner et al.*, 2003] with T63 ($\sim 1.9^\circ$) horizontal resolution, the ocean model MPI-OM [*Marsland et al.*, 2003] with a bipolar grid of 1.5° resolution, and the ocean biogeochemistry model HAMOCC [*Wetzel et al.*, 2005]. The land surface scheme is represented by the JSBACH model, which is based on the biosphere model, BETHY [*Knorr*, 2000]. General data preprocessing includes the fitting of the horizontal resolution of the model output data (T63) to that of the reanalysis data ($2.5 \times 2.5^\circ$). To assess the response of the small-scale predictand to changes of the large-scale predictors of the historical runs and the RCP4.5 scenario runs, the model data of the 700 hPa geopotential height fields are projected in each case onto the existing PCs of the observational period to obtain new predictor time series [for a detailed description of this approach, see e.g., *von Storch and Zwiers*, 1999; *Hertig*, 2004]. This is done for the seasonal mean large-scale atmospheric regimes obtained through S-mode PCA and for the station-specific daily circulation patterns derived from T-mode PCA.

3. Statistical Downscaling Approach

[15] For a schematic overview of the statistical downscaling approach, see Figure 3.

3.1. Modeling of Daily Precipitation in the Mediterranean Area

[16] The probability distribution of daily precipitation is highly skewed and does not have the constant variance properties required for multiple linear regression based on normal distribution of the predictand. Furthermore, in order to deal with zero values included in daily precipitation data, the modeling of precipitation is commonly split into a model for the occurrence and another one for the amount [see e.g., *Wilby et al.*, 2003]. However, in the present study, daily precipitation data are treated as if they had a Poisson distribution, consequently allowing for the handling of the data with one model. Similar to the approach of *Schoof and Pryor* [2001] to down-scale precipitation in Indianapolis, daily precipitation in the

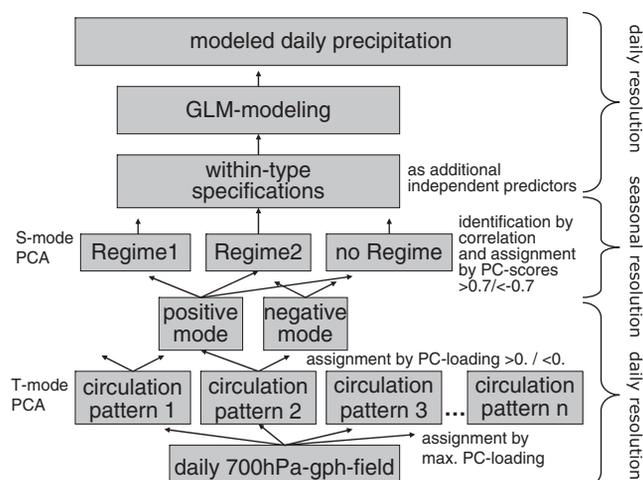


Figure 3. Schematic overview of the statistical downscaling approach.

Mediterranean area is modeled within the context of generalized linear models (GLMs) where the predictand is no longer assumed to be normally distributed. In the present study, the Poisson regression model has been used [for details on GLMs, see e.g., *McCullagh and Nelder*, 1989]. In order to predict the large number of zero values, the Poisson regression models mostly have small coefficients for the predictors. Thus, the downscaling method applied tends to underestimate precipitation events of large magnitude. To account for this, overdispersion is introduced to the models, allowing for a larger variance compared to the regular Poisson distribution where the variance is just proportional to the mean. The canonical log link function is applied to the mean of the response and the resulting model is fitted by the method of quaslikelihood. Figure 4 shows, for the example of daily winter precipitation at Be'er Scheva (Israel), a quantile–quantile plot in order to compare the observed and statistically modeled precipitation distributions. Additionally, the plot contains the 90% and 95% percentiles of observed (red symbols) and modeled precipitation (blue symbols). Figure 4 illustrates that in the modeled time series, there is a smaller number of dry days and precipitation is slightly overestimated for precipitation amounts approximately below the 90% percentile. Above this threshold, an underestimation becomes apparent. Overall, there is a reasonable agreement between observed and modeled precipitation, allowing for the modeling of mean daily precipitation amounts. However, it should also be pointed out that the applied GLM is not appropriate for modeling extreme precipitation. Rather, an extreme value distribution has to be adopted for this purpose.

[17] Here, a combined circulation-based and transfer function-based approach is used to assess daily precipitation. Based on the daily circulation patterns in station-specific domains obtained from T-mode PCA of daily 700 hPa geopotential height fields, daily precipitation at the Mediterranean stations is modeled within each season. First, each circulation pattern is checked for its relationship with the large-scale atmospheric regimes received from S-mode PCA of seasonal mean 700 hPa geopotential heights. This is done via correlation analysis of

circulation pattern time series (T-mode PC loadings) and large-scale regime time series (S-mode PC scores) on a seasonal basis. For highly significant correlations (99.9% level), composites are calculated for days with the simultaneous occurrence of a specific mode (positive or negative) of a circulation pattern and its related large-scale regime, and for days with the occurrence of this circulation pattern uncoupled to the appearance of the regime. Then, GLMs are established to assess the precipitation for each composite of the specific circulation pattern mode. In case of no correlation between a specific circulation pattern mode and a large-scale regime, no composite is calculated and just one GLM is used to model precipitation for this circulation pattern mode. As predictors in the GLMs, the time series of the circulation patterns and the standardized time series of the within-type characteristics (relative humidity, CIN, and u-wind and v-wind components) are taken. Within this statistical downscaling approach, the prevailing circulation pattern mode at a specific day of the observational time period is determined by taking the daily maximum PC loading as a decision criterion. Within verification, for the historical runs, and for the future projections the 700 hPa geopotential height fields are projected onto the existing PCs to obtain new predictor variables, and again, the daily maximum PC loading is used to select the prevailing circulation pattern at a specific day.

3.2. A Novel Approach to Validation

[18] Within the context of calibrating and verifying statistical models, validation measures have to be used to assess the quality of the statistical models. Simple measures of performance that can be applied to all different types of downscaling models are bias and mean squared error, which can be used for error assessment of the distributional mean and the variance, respectively. For the validation of time series, the correlation between observed and statistically modeled time series is in widespread use. In case of distributional assessments, statistical tests, quantile plots, and probability scores are common measures of model performance. For a more comprehensive overview on validation measures, see *Maraun et al.* [2010]. The approach introduced in the present study is based on bootstrap confidence intervals and can be adapted to various kinds of statistical downscaling approaches; that is, it does not make any assumptions concerning the underlying distributions and can be used for a wide variety of validation measures.

[19] The approach consists of various steps: first the whole time period available for a precipitation station is split into running 31 year subperiods; that is, 31 year periods are used each shifted by 1 year. When the end of the whole time series is reached, years from the beginning of the time series are successively included in order to avoid a more frequent inclusion of years located in the middle of the time series. Then, for each 31 year period, a statistical model is calibrated using the statistical downscaling approach presented in section 3.1. In each case, the years outside the calibration period are taken to validate the statistical model. Within the context of judging model performance, the errors related to the moments are examined; that is, for analyzing the first-order moment (mean of a distribution) the bias is considered. Thus, from the statistical downscaling results, the bias of statistically modeled daily precipitation and its 99% bootstrap confidence interval limits (using 1000 iterations of bootstrapping) are calculated for each calibration and each verification period.

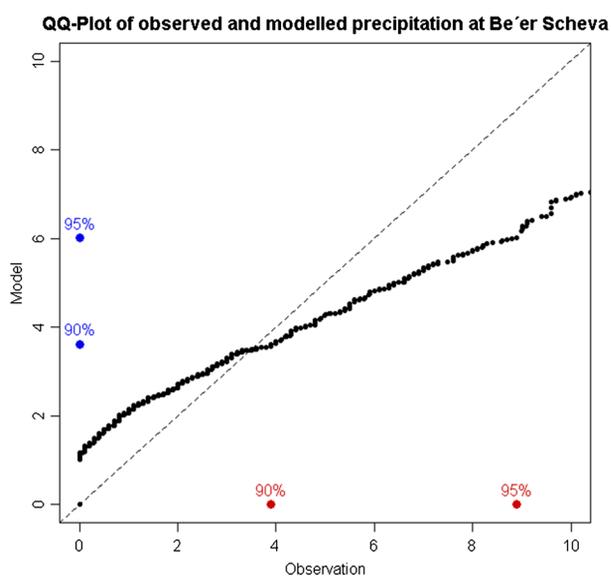


Figure 4. Q–Q plot comparing observed and modeled precipitation at Be'er Scheva (Israel) in winter (December–February).

Bias is calculated as the modeled minus the observed daily precipitation. Also, the mean bias is calculated from the biases of all calibration and verification periods, respectively, to obtain a measure of the overall model performance. To decide whether the statistical model for a station is appropriate in general, the mean biases in calibration and verification are checked for their location within the potential spread of the observational precipitation mean. This is done by looking at the overlap of the confidence interval of the mean bias in calibration and verification, respectively, with the 99% bootstrap confidence interval of the observed precipitation mean. Then, the 99% bootstrap confidence intervals of the bias calculated for each calibration/verification period are taken to judge model performance in the specific periods with regard to the overall model performance. This is done by looking at the overlap of the corresponding confidence intervals. If the confidence interval of a specific 31 year period does not overlap with the confidence interval of the mean bias, model performance is significantly different in this time period and nonstationarity of the model performance becomes evident. However, if just one calibration/validation period shows a nonstationary behavior, one might erroneously reject the null hypothesis of no significant difference. Therefore, nonstationarity is only stated if the confidence intervals of several successive calibration or validation periods lie outside the confidence interval of the mean bias. Subsequently, within the scope of the statistical assessments for the 21st century, for every precipitation station, a statistical model showing stationary model behavior in calibration and verification is used. In case of nonstationarities for a particular station, statistical models which describe these nonstationary relationships are also included. This leads to the application of one statistical model in case of overall stationary model behavior at a specific station, or of a statistical model ensemble in case of nonstationarities.

[20] In summary, the general appropriateness of a statistical model is considered by checking that the overall mean model error is not greater than the potential spread of the observations. Subsequently, the analysis of nonstationarities is based on significant departures of model errors in 31 year subperiods compared to the overall mean model error. Thus, nonstationarity analysis refers to the temporally varying statistical model performances under the acknowledgment of a specific model-inherent error. In the end, the approach leads to a substantiated selection of specific statistical downscaling models for the future assessments. Results of the nonstationarity analyses for mean daily precipitation at various Mediterranean sites are presented in section 4.2, the reasons for nonstationarities arising are discussed in section 4.3, and the application of the approach to Mediterranean precipitation in the 21st century is shown in section 4.4.

4. Results

4.1. Predictor Characteristics in Reanalysis and GCM Runs

[21] Because the statistical downscaling approach applied here strongly rests on large-scale atmospheric circulation, it is necessary to assess the capability of the GCM to reproduce the dominant large-scale circulation regimes. Figure 2a shows the large-scale atmospheric regimes for winter in the period 1950 to 2010 as obtained from S-mode PCA of seasonal mean 700 hPa geopotential height fields of the reanalysis data. It becomes evident that the four regimes extracted generally

represent the well-known modes of variation of the North Atlantic–European sector in winter. This is also reflected in the high correlation coefficients (0.6 up to 0.88, see Figure 2a) with Climate Prediction Center indices describing the NAO, EA, SCAND, and EA/WR Pattern. Figure 2b shows the corresponding large-scale atmospheric regimes from MPI-ESM-LR run 1. The patterns are sorted in this figure to match the reanalysis patterns. It can be seen that the overall pattern structures are quite similar to those in the reanalysis. But some differences also appear. Most notably, there are changed amounts of explained variances of the regimes related to the EA Pattern and the EA/WR Pattern. There are also some differences in all patterns regarding the location and strength of the main centers of variation. Some seasonal differences become apparent, being strongest in summer and autumn (data not shown). Examining the three MPI-ESM-LR runs shows that there is generally a high consistency of the runs, although with some differences regarding the amount of explained variance of the PCs and some spatial shifts and/or differences in intensity of the centers of variation. It should be noted that the GCM patterns presented in this section are only used to assess their correspondence to the reanalysis patterns; they are not used directly within the statistical downscaling approach. To assess the response of precipitation to changes in the large-scale predictors of the GCM runs, the model data are projected in each case onto the existing PCs of the observational period to obtain new predictor time series. However, it should be pointed out that the highlighted differences between reanalysis and GCM data will contribute to the uncertainties of the downscaling results.

[22] Regarding the variables used to describe the within-type characteristics of circulation patterns, it has been found that relative humidity, CIN, and the horizontal wind components show nonsystematic, pattern-specific differences between reanalysis and model data. A consistent pattern can only be seen in mostly higher values of 700 hPa relative humidity and stronger CIN in the GCM data compared to reanalysis.

[23] Notwithstanding the deficiencies of the MPI-ESM-LR model to capture all aspects of natural variability correctly [see also e.g., Brands *et al.*, 2012], the pressure-related variables in particular are regarded as useful predictors to assess regional climate change. Yet within the discussion of the results, the shortcomings of the GCM have to be kept in mind.

4.2. Running Calibration Analysis

[24] To illustrate the results of the running calibration analyses, an example is taken from the eastern Mediterranean area in the winter season: Figure 5 visualizes the results for the station at Be'er Scheva (Israel). The time series covers the years 1950 to 2004. Shown is the mean bias of statistically modeled daily precipitation and its 99% bootstrap confidence interval limits, calculated from the biases of all calibration periods (Figure 5a) and verification periods (Figure 5b). Mean bias of the calibration periods amounts to -0.18 mm/d. The narrow confidence interval (-0.17 mm/d to -0.19 mm/d) indicates that this systematic underestimation of precipitation occurs in all calibration periods with similar magnitude. For the verification periods, a mean bias of only -0.08 mm/d is calculated, but with a wider confidence interval (-0.17 mm/d to 0.0 mm/d) due to a larger spread of the biases in the verification periods. Both confidence intervals overlap within the potential range of the observed precipitation mean

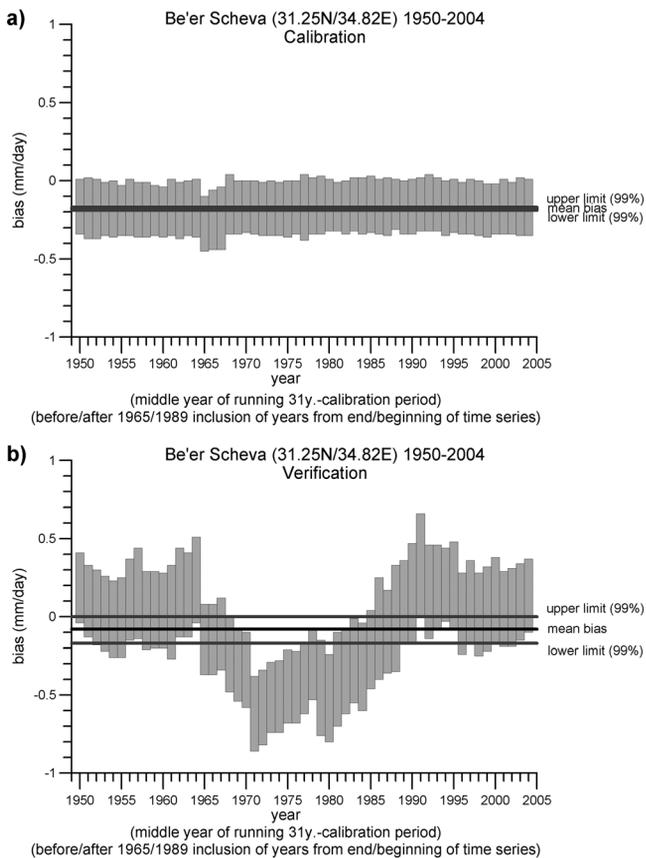


Figure 5. Bootstrap confidence intervals (99% significance level) of the bias of statistically modeled daily winter precipitation at Be'er Scheva in each 31 year subperiod (gray bars), mean bias (black line), and its 99% bootstrap confidence interval limits (gray lines), calculated from the biases of all calibration periods (a) and verification periods (b).

(1.37 ± 0.2 mm/d), justifying the use of the statistical model in general. Now, the 99% bootstrap confidence intervals of the bias calculated for each 31 year calibration/verification period (gray bars in Figure 5) are taken to judge model performance in the specific periods with regard to the overall model performance. This is done by looking at the overlap of the confidence intervals. In case there is no overlap, the bias of a particular period is significantly different from the mean bias and a nonstationarity is postulated. For the example in Figure 5, nonstationarities occur in series within the validation of the models from the calibration periods 1956–1986 to 1961–1991 (see gray bars at the corresponding central years 1971–1976). This implies that predictors–precipitation relationships established between the mid-1950s and at the beginning of the 1990s are less adequate to describe conditions outside this time period, resulting in a negative bias of modeled precipitation in the corresponding verification years. The causes for the detected nonstationarities are discussed in section 4.3.

[25] Looking at the results for the other analyzed precipitation stations in the Mediterranean area reveals a diverse picture of nonstationarities. From the 22 precipitation time series considered in the winter season, one station's statistical model does not have an adequate overall model performance, 14 stations reveal nonstationarities, and 7 stations get

stationary statistical models within the station-specific time periods considered. In spring, precipitation models do not pass the overall performance measure at two stations, models for nine stations show nonstationarities, and models for seven stations are stationary. In the summer season, statistical models at eight stations fail to pass the overall performance criterion due to difficulties in representing the very low mean precipitation amount and the very high precipitation variability at these stations during summer. For the remaining 15 stations, 8 showed nonstationary model behavior whereas the predictors–predictand relationships remained stationary at 7 stations. In contrast, during autumn, precipitation has a nonstationary connection to the predictors at only 6 stations, whereas for 16 stations, the statistical models are stationary. In summary, only about one third of the stations exhibit nonstationarities in autumn. In spring and summer, approximately half of the stations are affected by nonstationarities; in the winter season, up to approximately two thirds are affected. Thus, winter, the time of the year when the Mediterranean area is under the increased influence of the midlatitude circulation, represents the season when most nonstationarities occur.

[26] From the results obtained for the various stations, a generalized picture of major nonstationarities in different regions of the Mediterranean area can be drawn. For the winter season, the nonstationarity at Be'er Scheva described above is also visible for other eastern Mediterranean stations in Israel to Cyprus. For the western Mediterranean area, a remarkable change occurred between the period from 1950 to the mid-1980s and the years afterward, affecting stations from the Atlantic coast of Portugal, over central-northern to northeastern Spain. For the northern Mediterranean area (northern Italy and the east coast of the Adriatic Sea), nonstationarities in the predictors–precipitation relationships could be seen between the periods before and after the end of the 1970s. In spring, major nonstationarities occurred over many parts of the Iberian Peninsula for the years before and after the mid-1960s. Over the northern Mediterranean area from France to Greece, the beginning of the 1980s marked a significant changing point. In the summer season, precipitation over the whole Iberian Peninsula was affected by nonstationarities for years before the early to mid-1970s compared to the subsequent periods. In autumn, no temporal and regional structure of the observed nonstationarities came to the fore.

4.3. Analysis of Nonstationarities

[27] In order to show the underlying sources for the nonstationarities assessed in the previous section, we return to the example of winter precipitation at Be'er Scheva. The dominant circulation patterns (first two PCs with explained variances of approximately 43% and 31%, respectively) in the domain tailored for precipitation at Be'er Scheva are shown in Figure 6. The positive mode of PC1 yields anticyclonic conditions over the eastern Mediterranean area, and low pressure toward Central Europe (Figure 6a). This mode is connected to the large-scale atmospheric regime representing the negative phase of the NAO (correlation coefficient, 0.61) as well as to the regime that resembles the negative EA/WR Pattern (correlation coefficient, 0.47). Thus, this circulation pattern is primarily connected to these large-scale regimes, but the pattern can also evolve from other large-scale circulation conditions, e.g., for all winter seasons from 1950 to 2004,

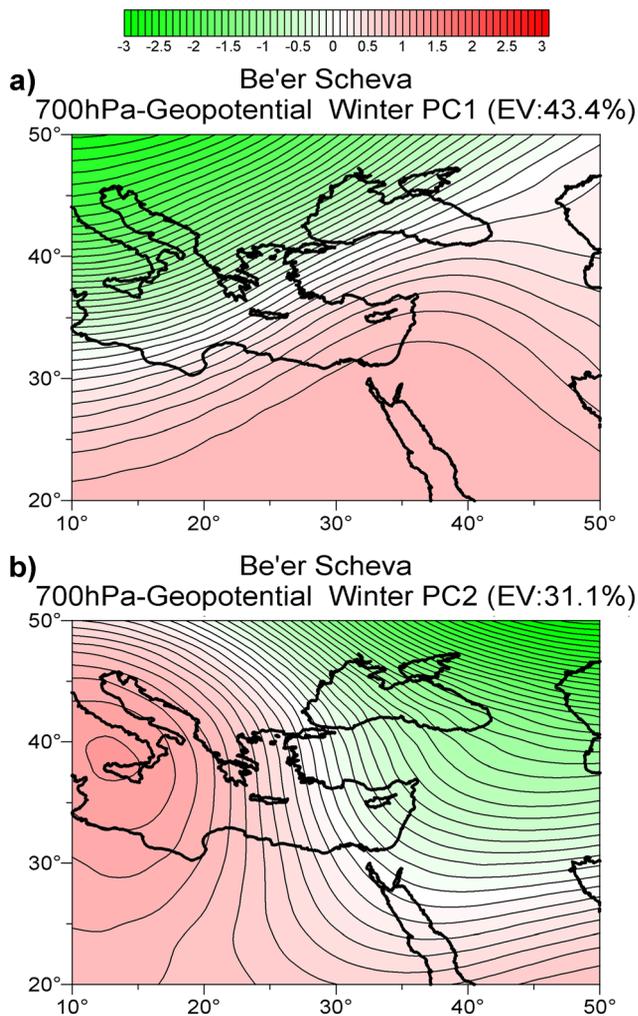


Figure 6. Circulation patterns (first two T-mode PCs, 700 hPa level) in the domain tailored for precipitation at Be'er Scheva.

the positive mode of PC1 occurred at 2756 days (55.5% of all winter days), but of these, only 872 days (31.6%) were connected to the negative NAO, and 394 days (14.3%) to the negative EA/WR Pattern. The positive mode of PC2 shows high pressure over the central Mediterranean area, whereas the eastern Mediterranean area is influenced by a low pressure system centered over eastern Europe/Russia (Figure 6b). This leads to a northwestern cyclonic flow into the eastern Mediterranean area. The positive mode of PC2 is correlated with the large-scale atmospheric regime representing the positive EA Pattern (correlation coefficient, 0.46). For all winter seasons from 1950 to 2004, the positive mode of PC2 occurred on 1739 days (35% of all days), with 388 days (22.3%) connected to the positive EA Pattern. The positive mode of PC2 is predominantly responsible for the generation of precipitation days at Be'er Scheva: 565 precipitation days with an average amount of 2.82 mm/d are associated with this circulation pattern (out of a total of 872 winter precipitation days in 1950–2004).

[28] Because the period from the mid-1950s to the early 1990s was identified as a period of changed relationships of the large-scale predictors with precipitation at Be'er Scheva (see section 4.2), the calibration period 1960 to 1990 is taken

to examine the nonstationary statistical model performance. Concerning the overall circulation pattern frequency and its relation to large-scale regimes, it becomes evident that the precipitation producing positive mode of PC2 occurs more frequently (36.5% of all winter days in 1960–1990) compared to the other years (28.6%), but it is less connected (20.4% compared to 24.5%) to the large-scale atmospheric regime related to the positive EA Pattern. Overall, there was a higher amount of mean daily rainfall from 1960 to 1990 (1.4 mm/d) compared to the years before and after this period (1.25 mm/d).

[29] Our approach derives a precipitation transfer function for each atmospheric regime composite of a circulation pattern. Inspection of the GLM predictors and their regression coefficients reveals that the causes for the nonstationary model performance has to be sought in differing relationships of precipitation with specific circulation patterns and their within-type characteristics. In the period 1960 to 1990, the GLMs of days with positive mode of PC1 yield no differences regarding the chosen predictors and their regression coefficients compared to GLMs of years outside this period. Even though there is a much stronger (weaker) connection of positive PC1 and negative NAO (EA/WR) in 1960 to 1990 compared to the years outside this period, the circulation pattern–precipitation relationships and the within-type characteristics of these configurations are stationary. In contrast, the GLMs of days with positive mode of PC2 associated with the positive EA Pattern show significantly lower regression coefficients compared to the ones from GLMs of years outside 1960 to 1990. Moreover, the GLMs of days with positive mode of PC2 not connected to the EA Pattern show much higher regression coefficients in the period 1960 to 1990. This implies that under the simultaneous occurrence of the positive mode of PC2 and the positive EA Pattern, a weaker signal from the predictors sufficed to induce precipitation from 1960 to 1990, partly due to the higher relative humidity and a stronger westerly wind component associated with this configuration. In contrast, for the positive mode of PC2, which is not connected to the EA Pattern, a stronger signal is required in 1960 to 1990. The varying strength of the predictors–precipitation relationships causes (enhanced by the changed frequencies of occurrence) an underestimation of precipitation in the periods outside the calibration.

[30] Other nonstationarities have been addressed for precipitation at various Mediterranean regions in different seasons of the year (see section 4.2). In this context, winter was identified as the season when most nonstationarities occurred. Thus, further exemplifications will continue to focus on this season. A detailed analysis of the wintertime nonstationarities reveals that those over the western Mediterranean area (from the Atlantic coast of Portugal over central-northern to northeastern Spain) from the period 1950 to the mid-1980s compared to the years afterward are caused by changes of the positive mode of the dominant circulation pattern for this region. The pattern shows high pressure over the western Mediterranean Sea and low pressure over the North Atlantic area, inducing a southwestern flow into the target area. The pattern has a strong connection to the atmospheric regime resembling the positive phase of the EA Pattern (correlation coefficient, 0.76). In the period before the mid-1980s, the positive mode of the circulation pattern occurs mostly uncoupled to the positive EA Pattern, and it is generally characterized by stronger

cyclonicity with high values of relative humidity, low CIN, and strong westerly flow. Consequently, in this period, a weaker signal from the predictors conforms to model precipitation in the GLMs under these pattern characteristics. The negative mode of a circulation pattern, featuring low pressure over the northern Iberian Peninsula and the adjacent North Atlantic Ocean, is also strongly related to the generation of precipitation over the northern half of the Iberian Peninsula. This circulation pattern strongly correlates to the atmospheric regime associated with the negative phase of the NAO (correlation coefficient, 0.75). In the years before the mid-1980s, the circulation pattern very frequently occurred with simultaneously negative phases of the NAO and was associated with much higher values of relative humidity, lower CIN, and stronger wind components. Thus, in the GLMs, rather small regression coefficients suffice to reliably model precipitation. When transferring the relationships of these two main precipitation-generating circulation patterns to the period after the mid-1980s, a major nonstationarity of the statistical model performance arises because the link provided is too weak to adequately model precipitation in the later periods, resulting in a negative bias of modeled precipitation even though observed precipitation was already lower in the later period.

[31] For the northern Mediterranean area during winter, nonstationarities in the predictors–precipitation relationships can be seen between the periods before and after the end of the 1970s. Despite higher values of the within-type characteristics of various circulation patterns, a smaller precipitation amount is connected with the occurrence of these circulation patterns in the period before the end of the 1970s, requiring only relatively small regression coefficients in the GLMs. But with the transference of these relationships to the period after the 1970s, the predictor signals on precipitation are too weak because, in this period, the circulation patterns have generally lower values of the within-type characteristics, whereas an increase of precipitation is associated with these patterns in the later period.

[32] In summary, it should be stressed that frequency changes of the circulation patterns alone are not the cause of nonstationarity in the statistical model performance because changing frequencies are accommodated by the statistical downscaling approach. In fact, the assessed nonstationarities are due to

varying predictors–precipitation relationships of specific circulation patterns. In this regard, frequency changes can damp or increase the effects of the nonstationary relationships.

4.4. Assessments for the 21st Century

[33] The identification and analysis of nonstationarities in the predictors–precipitation relationships lead to a substantiated selection of specific statistical downscaling models for future assessments. In general, for every precipitation station, a statistical model showing stationary performance in calibration and verification periods is used. In case nonstationarities have been detected for a particular station, additional statistical models are included that describe these nonstationary relationships. This leads to the application of one statistical model (in case of an overall stationary model behavior at a specific station) or of a statistical model ensemble (which is composed of two to three statistical ensemble members, depending on the number of identified nonstationarities for a specific station). For the example of winter precipitation at Be'er Scheva, two statistical downscaling models are used for the assessment of precipitation changes until the end of the 21st century. The calibration period, which comprises the years 1950 to 1961 and 1986 to 2004, is taken for a statistical model describing stationary conditions, whereas the model from the calibration period 1956 to 1986 is used to include nonstationary predictors–precipitation relationships. Figure 7 shows that, for the example of Be'er Scheva, the temporal evolution of mean daily precipitation in winter for the period 1950 to 2099 in terms of the polynomial trend of each individual assessment and of the overall ensemble mean. The graph comprises two statistical downscaling models using three MPI-ESM-LR runs under historical (1950–2005) and RCP4.5 scenario (2006–2099) conditions. It can be seen that the overall projected trend of daily precipitation is slightly decreasing until the 2020s and increases afterward, with an overall increase of approximately 0.2 mm/d. When looking at the spreads caused by the application of different statistical models and of different GCM runs, it becomes evident that the uncertainty range, which arises from the statistical models, is about twice as large as the range resulting from the application of different GCM runs. The distinct relationships of the predictors with precipitation, defined by the regression coefficients in a specific statistical

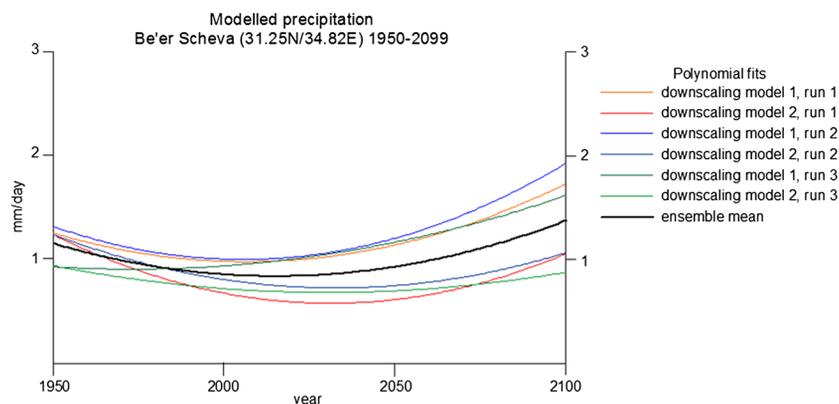


Figure 7. Temporal evolution of mean daily precipitation at Be'er Scheva in winter for the period 1950 to 2099 resulting from two statistical downscaling models using three MPI-ESM-LR runs under historical (1950–2005) and RCP4.5 scenario (2006–2099) conditions. Shown are the polynomial fits of each individual assessment and of the overall ensemble mean.

model, have a strong impact on the future evolution of precipitation. However, the substantiated selection of statistical models leads to an incorporation of the whole range of observed predictors–precipitation relationships, and gives a good estimate of the overall uncertainty arising from natural variations in the circulation–climate relationships. Of course, these uncertainties do not stand alone, but have to be integrated in the large uncertainty range inherent to climate modeling. This aspect will be picked up again in the discussions of section 5.

[34] Figure 8 illustrates the ensemble mean changes (ensemble mean of statistical models and GCM runs) of mean daily precipitation across all included stations for the future period 2070 to 2099 in relation to the historical model period 1961 to 1990 under RCP4.5 scenario assumptions. In addition, the changes in percentages in relation to the mean of the period 1961 to 1990 and the significance of the changes (95% level, U test) are shown. In winter, significant increases become apparent over large parts of the western and northern Mediterranean area, whereas the eastern and southern parts are mostly affected by precipitation reductions until the end of the 21st century. In spring, summer, and autumn, decreases of precipitation clearly dominate over the entire Mediterranean area, with the strongest reductions over the western and southern Iberian Peninsula in autumn. A noticeable exception occurs for some southern and central regions of the Iberian Peninsula in summer with projected precipitation increases. The obtained overall pattern of station-based daily precipitation changes under RCP4.5 scenario assumptions is mostly in agreement with the statistical assessment results for grid-based seasonal precipitation totals under A1B scenario assumptions [Hertig *et al.*, 2012] and also with the general coarse-grid pattern of changes as simulated by various GCMs [Giorgi and Lionello, 2008].

5. Discussion and Conclusions

[35] In the present study, daily precipitation in the Mediterranean area was assessed by a combined circulation-based and transfer function–based statistical downscaling approach. In this context, large-scale seasonal atmospheric regimes, synoptic-scale daily circulation patterns, as well as their within-type characteristics have been related to daily station-based precipitation. Information on the large-scale atmospheric state has been incorporated by composites of the related smaller-scale circulation patterns. For each composite and its within-type characteristics, a GLM, assuming a Poisson process with the canonical link log, has been derived to assess precipitation. It has been found that the specified downscaling approach is suitable to assess mean daily precipitation at various stations across the Mediterranean area. However, for the assessment of extreme precipitation, other approaches should be considered; that is, an extreme value distribution or a mixture model should be applied. Thus, the Poisson–GLM could be extended to a mixture model by combining it with a Gamma–GLM, or a generalized Pareto distribution could be applied.

[36] We introduced a novel approach to validation under the explicit consideration of nonstationarities in the predictors–predictand relationships. The validation approach suggested in the present study does not depend on a specific statistical method. The principle of using running calibration analysis and bootstrap confidence intervals should be applicable to a

wide range of downscaling techniques, as well as with those dealing with extremes. The approach is based on the results from running calibration periods. The whole study period is split into 31 year subperiods, each shifted by 1 year; for each subperiod, a statistical model is calibrated and subsequently validated in the years outside of the calibration period. Nonstationarities are then detected by comparing the overall model performance (derived by averaging the performances of all subperiods) to the model performance of an individual subperiod. The (non)overlaps of the bootstrap confidence interval of the mean model performance and the bootstrap confidence intervals of the individual model errors are used to identify (non)stationary model performance. The 99% significance level has been applied for the confidence intervals. In general, the length of the calibration periods and the significance level for the bootstrap confidence intervals might be adjusted individually. Note, however, that the confidence interval limits strongly depend on the sample size; that is, reducing the length of the subperiods widens the confidence intervals.

[37] In the present study, the specified approach to validation has been demonstrated for mean precipitation using the bias to assess model skill. However, the method can also be generalized to other validation measures. As an example, the results of a running calibration analysis with the correlation coefficient between observed and modeled precipitation as performance measure is shown in Figure 9 for Barcelos (Portugal) in winter. In this example, the correlation coefficients trace the same nonstationarity that has been found for the bias. In Figure 9, the nonstationarity is expressed by significantly lower correlation coefficients in the verification years of the models, which are trained in the years from 1950 to the mid-1980s (see gray bars at the years 1966 to 1971 in Figure 9b, see also sections 4.2 and 4.3 for further details on the nonstationarity). It would also be possible to apply the validation approach to the mean squared error to judge model performance regarding precipitation variance. However, this would require that the modeling of precipitation variances within statistical downscaling approaches becomes much more adequate than it is today.

[38] Within the analysis of mean daily precipitation in the Mediterranean area, nonstationarities occurred within the relationships of precipitation with synoptic-scale circulation patterns and their within-type characteristics. The circulation patterns themselves are connected to the midlatitude variability, being strongest in the winter season. The midlatitude variability can be described by well-known atmospheric regimes, i.e., the NAO, EA, SCAND, and EA/WR Pattern [a detailed description of the relations of these modes to Mediterranean climate is given, e.g., by Trigo *et al.*, 2006]. It was shown that frequency changes of the synoptic-scale circulation patterns alone are not the cause of nonstationarity in the statistical model performance because changing frequencies are accommodated by the statistical downscaling approach. Rather, the assessed nonstationarities are caused by varying predictors–precipitation relationships of specific circulation patterns. In this regard, frequency changes can damp or increase the effects of the nonstationary relationships. Reasons for varying relationships within specific circulation patterns and hence for nonstationarities of the statistical model performance might be associated with changing intensities and positions of the centers of variation of the large-scale atmospheric regimes. For example, Beranová and Huth [2008], using a running correlation analysis of 500 hPa circulation modes and temperature and precipitation across Europe during winter

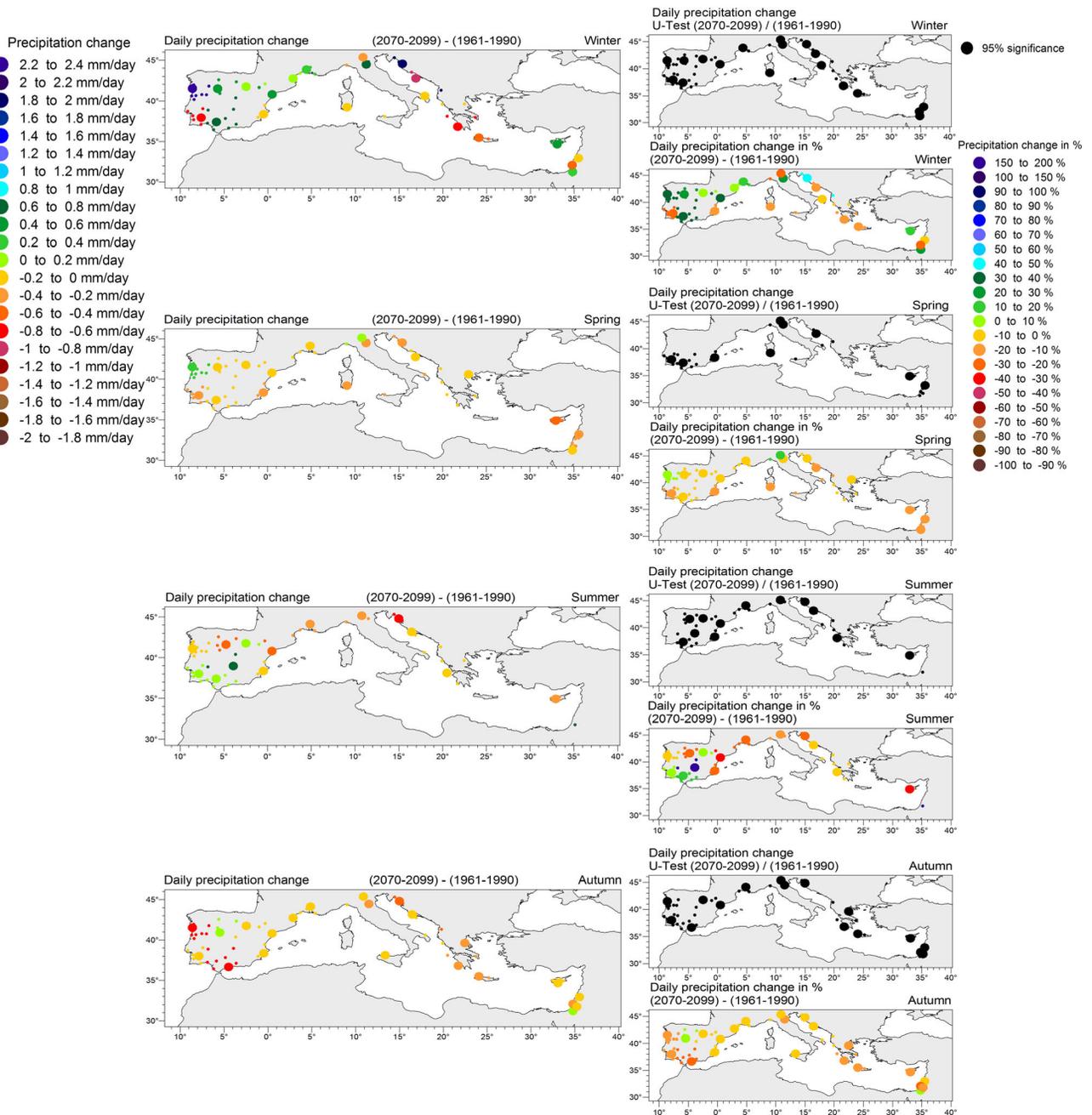


Figure 8. Ensemble changes of mean daily precipitation in millimeters per day for the future period 2070 to 2099 in relation to the historical model period 1961 to 1990 under RCP4.5 scenario assumptions. In addition, the changes in percentages in relation to the mean of the period 1961 to 1990, and the significance of the changes (95% level, U test) are shown. Results are based on statistical downscaling assessments using GLMs as downscaling technique and predictors from three MPI-ESM-LR runs. From top to bottom: winter, spring, summer, autumn. Big symbols, representative stations; small symbols, associated stations.

for the period 1958 to 1998, found that a decrease in NAO correlations at the precipitation stations south of 45°N since 1985 could be attributed to an enhanced anticyclonicity caused by an eastward shift of the southern (anticyclonic) center of the NAO leading to a lowering of the correlations. In the same study, it was found that the largest change in the EA Pattern occurred during the period 1986 to 1998, with an intensification of the EA Pattern together with an extension of its northern center toward Scandinavia. This results in increasingly negative correlations over the western and northern

Mediterranean area. For the SCAND Pattern [EU1 Pattern in *Beranová and Huth, 2008*] the center over Spain underwent the largest changes in its shape and position, becoming strongest and most dominant in the latest period analyzed (1986–1998). This is associated with a much more westerly than northerly flow toward southern Europe in its negative mode. *Fealy and Sweeney [2005]* identify in an analysis of Scandinavian glaciers a significant change point at the end of the 1980s with major changes in the large-scale atmospheric variability of the North Atlantic region and

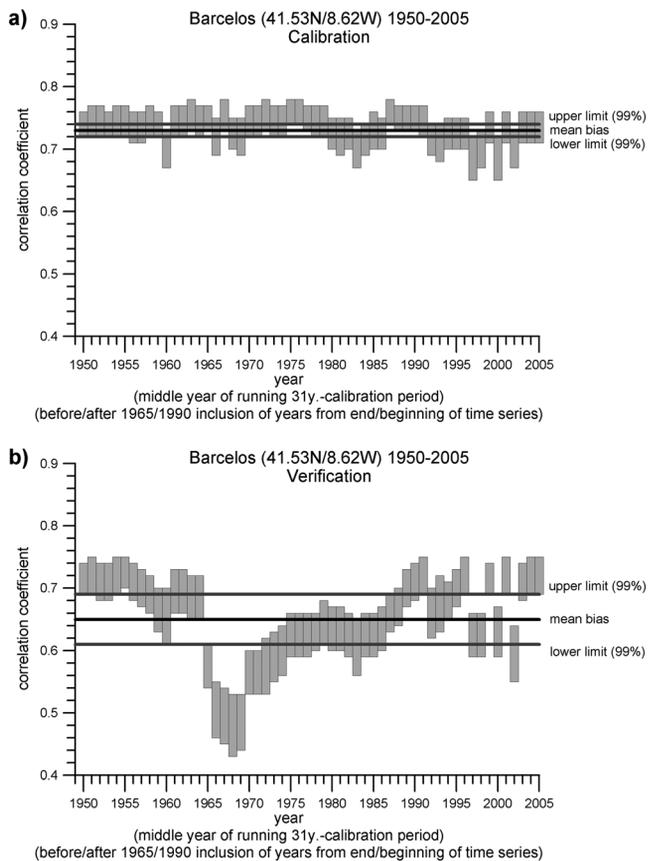


Figure 9. Bootstrap confidence intervals (99% significance level) of the correlation coefficient between statistically modeled and observed daily precipitation in winter at Barcelos in each 31 year subperiod (gray bars), mean correlation coefficient (black line), and its 99% bootstrap confidence interval limits (gray lines) calculated from the correlation coefficients of all calibration periods (a) and verification periods (b).

associated temperature and precipitation characteristics. The authors conclude that this change point may be part of a larger-scale global atmospheric variability change. Overall, the reasons for nonstationarities coming from changes in the large-scale atmospheric circulation can propagate to smaller scales. Thus, the station-specific circulation patterns and their within-type characteristics vary partly parallel to changes of the large-scale atmospheric regimes. However, the smaller-scale circulation patterns also feature their own distinct temporal variations.

[39] In the context of assessing future precipitation changes under increased greenhouse warming conditions, the identification and analysis of nonstationarities in the predictors–precipitation relationships leads to a substantiated selection of specific statistical downscaling models for future assessments. This is of considerable importance because the nonstationarities of the predictors–predictand relationships found under current climate conditions give rise to the assumption of future nonstationarities. Thus, the downscaling approach presented here also supports the understanding of uncertainties and leads to improved statistical downscaling assessments. Generally, uncertainties in climate modeling can be roughly grouped into three areas. The first is related to uncertainties arising from external influences

(i.e., unknown future anthropogenic greenhouse gas emissions). The second aspect relates to the limited knowledge of the climate system (e.g., carbon sinks, climate sensitivities, feedbacks, and tipping elements). The third aspect addresses uncertainties inherent in global and regional climate models. These are mainly associated with the spatial and temporal resolution, the issues of discretization and parametrizations, unknown initial conditions, and the degree of reproduction of several important variables such as sea surface temperatures [Rowell, 2012], soil moisture [Dirmeyer *et al.*, 2006], carbon cycle, and stratospheric processes [Dai, 2006]. Rowell [2012], using perturbed physics, a multimodel, and a CMIP3 ensemble, assesses that over many of the midlatitude regions, internal atmospheric variability determines much of the variety in projections of 21st century local precipitation. However, for the midlatitude continental areas in summer deficits in modeling of atmospheric, land surface, and carbon cycle processes are of considerable importance. In the scope of regional to local assessments uncertainties inherent in downscaling approaches have to be addressed as well. In this context, nonstationarities, which arise from varying predictors–predictand relationships, as pointed out in the present study, have to be considered important contributors to uncertainty. Regarding the uncertainty range caused by the application of different statistical models and of different GCM runs, it becomes evident that the range that arises from taking nonstationarities in the observed predictors–predictand relationships into account is about twice as large as the range resulting from the application of different initial conditions in GCM runs. But besides the consideration of nonstationarities in observed predictors–predictand relationships, future work will also have to analyze the earth system model runs in more detail regarding this aspect.

[40] In the present study for daily precipitation in the Mediterranean area in winter, significant increases become apparent over large parts of the western and northern regions, whereas the eastern and southern parts are mostly affected by precipitation reductions using RCP4.5 scenario assumptions. In spring, summer, and autumn, decreases of precipitation until the end of the 21st century clearly dominate over the entire Mediterranean area. The strongest reductions can be seen over the western and southern Iberian Peninsula in autumn.

[41] Besides assessments of mean precipitation changes, reliable estimates of the probability of extreme precipitation changes are required for many applications such as environmental engineering and planning. However, nonstationarities also occur in downscaling assessments of extremes and might be especially pronounced. Thus, future work will concentrate on the projection of precipitation extremes under the explicit consideration of nonstationarities. Within this framework, the novel validation approach of the present study will be tested for its suitability to work on nonstationary conditions in the assessment of extremes.

[42] **Acknowledgments.** This project is funded by the German Research Foundation under contract HE 6186/2-1. The authors thank Jonathan Eden, Rowan Fealy, and two anonymous reviewers for providing useful comments.

References

Alexandersson, H. (1986), A homogeneity test applied to precipitation data, *J. Climatol.*, 6, 661–675.

- Barnston, A. G., and R. E. Livezey (1987), Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, *Mon. Wea. Review*, *115*, 1083–1126.
- Beckmann, B. R., and T. A. Buishand (2002), Statistical downscaling relationships for precipitation in the Netherlands and North Germany, *Int. J. Climatol.*, *22*, 15–32.
- Ben-Gai, T., A. Bitan, A. Manes, P. Alpert, and Y. Kushnir (2001), Temperature and surface pressure anomalies in Israel and the North Atlantic Oscillation, *Theor. Appl. Climatol.*, *69*, 171–177.
- Beranová, R., and R. Huth (2008), Time variations of the effects of circulation variability modes on European temperature and precipitation in winter, *Int. J. Climatol.*, *28*, 139–158.
- Brands, S., S. Herrera, J. Fernandez, and J. Gutierrez (2012), How well do CMIP5 Earth System Models simulate present climate conditions?, *Clim. Dyn.*, (in press).
- Busuioc, A., H. von Storch, and R. Schnur (1999), Verification of GCM-Generated Regional Seasonal Precipitation for Current Climate and of Statistical Downscaling Estimates under Changing Climate Conditions, *J. Climate*, *12*, 258–272.
- Cavazos, T., and B.C. Hewitson (2005), Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation, *Clim. Res.*, *28*, 95–107.
- Charles, S. P., B. C. Bates, P. H. Whetton, and J. P. Hughes (1999), Validation of downscaling models for changed climate conditions: case study of southwestern Australia, *Clim. Res.*, *12*, 1–14.
- Dai, A. (2006), Precipitation characteristics in eighteen coupled climate models, *J. Climate*, *19*, 4605–4630.
- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki (2006), The second global soil wetness project (GSWP-2): multi-model analysis and implications for our perception of the land surface, *Bull. Amer. Meteor. Soc.*, *87*, 1381–1397.
- Fealy, R., and J. Sweeney (2005), Detection of a possible change point in atmospheric variability in the North Atlantic and its effect on Scandinavian glacier mass balance, *Int. J. Climatol.*, *25*, 1819–1833.
- Giorgi, F., and P. Lionello (2008), Climate change projections for the Mediterranean area, *Glob. Planet. Change*, *63*, 90–104.
- Greatbatch, R. J., J. Lu, and K. A. Peterson (2004), Nonstationary impact of ENSO on Euro-Atlantic winter climate, *Geophys. Res. Lett.*, *31*, L02208, doi:10.1029/2003GL018542.
- Hertig, E., S. Seubert, A. Paxian, G. Vogt, H. Paeth, and J. Jacobet (2012), Changes of total versus extreme precipitation and dry periods until the end of the 21st century: statistical assessments for the Mediterranean area, *Theor. Appl. Climatol.*, DOI: 10.1007/s00704-012-0639-5.
- Hertig, E., and J. Jacobet (2008), Assessments of Mediterranean precipitation changes for the 21st century using statistical downscaling techniques, *Int. J. Climatol.*, *28*, 1025–1045.
- Hertig, E. (2004), Niederschlags- und Temperaturabschätzungen für den Mittelmeerraum unter anthropogen verstärktem Treibhauseffekt. PhD thesis published online at <http://opus.bibliothek.uni-wuerzburg.de/opus/volltexte/2004/874/>
- Hewitson, B. C., and R. G. Crane (2006), Consensus between GCM climate change projections with empirical downscaling: precipitation downscaling over South Africa, *Int. J. Climatol.*, *26*, 1315–1337.
- Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck (Eds.) (2003), *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, Geophys. Monogr. Ser., vol. 134, AGU, Washington, D. C., doi:10.1029/GM134.
- Jacobet, J., A. Dünkeloh and E. Hertig (2007), Mediterranean rainfall changes and their causes, in *Global Change: Enough Water for all?*, edited by J. Lozan, et al. 195–199, Wissenschaftliche Auswertungen, Hamburg.
- Kalnay, E., et al. (1996), The NCEP/NCAR 40-Year Reanalysis Project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.
- Kistler, R., et al. (2001), The NCEP/NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation, *Bull. Am. Meteorol. Soc.*, *82*, 247–268.
- Klein Tank, A. M. G., et al. (2002), Daily dataset of the 20th-century surface air temperature and precipitation series for the European climate assessment, *Int. J. Climatol.*, *22*, 1441–1453.
- Knorr, W. (2000), Annual and interannual CO₂ exchange of the terrestrial biosphere: Process based simulations and uncertainties, *Global Ecol. Biogeogr.*, *9*, 225–252.
- Kunstmann, H., A. Heckl, and A. Rimmer (2006), Physically based distributed hydrological modeling of the Upper Jordan catchment and investigation of effective model equations, *Adv. Geosci.*, *9*, 123–130.
- Maraun, D., et al. (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Rev. Geophys.*, *48*(3), RG3003, DOI 10.1029/2009RG000314.
- Mariotti, A., N. Zeng, and K.-M. Lau (2002), Euro-Mediterranean rainfall and ENSO - a seasonally varying relationship, *Geophys. Res. Lett.*, *29*, 1621, DOI 10.1029/2001GL014248.
- Marsland, S. J., H. Haak, J. H. Jungclaus, M. Latif, and F. Roeske (2003), The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates, *Ocean Model.*, *5*, 91–127.
- McCullagh, P., and J. A. Nelder (1989), *Generalized linear models, Monographs on Statistics and Applied Probability 37*, Chapman & Hall, London.
- Miller, A. J., D. R. Cayan, T. P. Barnett, N. E. Graham, and J. M. Oberhuber (1994), The 1976–77 climate shift of the Pacific Ocean, *Oceanogr.*, *7*, 21–26.
- Moberg, A., P. D. Jones, D. Lister, A. Walther, M. Brunet, J. Jacobet, L. V. Alexander, P. Della-Marta, J. Luterbacher, P. Yiou, et al. (2006), Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000, *J. Geophys. Res. Atmos.*, *111*, D22106, doi:10.1029/2006JD007103.
- Murphy, J. (1999), An Evaluation of Statistical and Dynamical Techniques for Downscaling Local Climate, *J. Climate*, *12*, 2256–2284.
- Myoung, B., Nielsen-Gammon, J. W. (2010), Sensitivity of monthly convective precipitation to environmental conditions, *J. Climate*, *23*, 166–188.
- Philipp A., P. M. Della-Marta, J. Jacobet, D. R. Fereday, P. D. Jones, A. Moberg, and H. Wanner (2007), Long term variability of daily North Atlantic-European pressure patterns since 1850 classified by simulated annealing clustering, *J. Climate*, *20*(16), 4065–4095.
- Preisendorfer, R. W. (1988), *Principal Component Analysis in Meteorology and Oceanography, Developments in Atmospheric Science 17*, Elsevier, Amsterdam.
- Quadrelli, R., V. Pavan, and F. Molteni (2001), Wintertime variability of Mediterranean precipitation and its links with large-scale circulation anomalies, *Clim. Dynam.*, *17*, 457–466.
- Roeckner, E., G. Baeuml, L. Bonaventura, R. Brokopf, M. Esch, S. Giorgetta, S. Hagemann, I. Kirchner, L. Kornbluh, E. Manzani, et al. (2003), The atmospheric general circulation model ECHAM5. Part I: Model description, *Max Planck Institute for Meteorology Report 349*, Hamburg.
- Rowell, D. P. (2012), Sources of uncertainty in future changes in local precipitation, *Clim. Dyn.*, *39*, 1929–1950.
- Schoof, J. T., and S. C. Pryor (2001), Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks, *Int. J. Climatol.*, *21*, 773–790.
- Trenberth, K. E. (1990), Recent observed interdecadal climate changes in the Northern Hemisphere, *Bull. Amer. Meteor. Soc.*, *71*, 993–998.
- Trigo, R., et al. (2006), Relations between variability in the Mediterranean region and mid-latitude variability, in *Mediterranean climate variability. Developments in Earth and Environmental Sciences, 4*, edited by Lionello, P., Malanotte-Rizzoli, P., Boscolo, R., pp. 180–226, Elsevier, Amsterdam.
- Van Vuuren, D. P., et al. (2011), The representative concentration pathways: an overview, *Clim. Chang.*, *109*, 5–31.
- Von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, pp. 484.
- Von Storch, H., E. Zorita, and U. Cubasch (1993), Downscaling of global climate estimates to regional scales: an application to Iberian rainfall in wintertime, *J. Climate*, *6*, 1161–1171.
- Vrac, M., M. L. Stein, K. Hayhoe, and X. Z. Liang (2007), A general method for validating statistical downscaling methods under future climate change, *Geophys. Res. Lett.*, *34*, L18,701, doi:10.1029/2007GL030,295.
- Wilby, R. L., O. L. Tomlinson, and C. W. Dawson (2003), Multi-site simulation of precipitation by conditional resampling, *Clim. Res.*, *23*(3), 183–194.
- Wetzel, P., A. Winguth, and E. Maier-Reimer (2005), Sea-to-air CO₂ fluxes from 1948 to 2003, *Global Biogeochem. Cycles*, *19*, GB2005, doi:10.1029/2004GB002339.
- Wijngaard, J. B., A. M. G. Klein Tank, and G. P. Können (2003), Homogeneity of 20th century European daily temperature and precipitation series, *Int. J. Climatol.*, *23*, 679–692.
- Wang, B. (1995), Interdecadal changes in El Niño onset in the last four decades, *J. Climate*, *8*, 267–285.
- Wilby, R. L. (1998), Statistical downscaling of daily precipitation using daily airflow and seasonal teleconnection indices, *Clim. Res.*, *10*, 163–178.
- Wilby, R. L., and T. M. L. Wigley (1997), Downscaling general circulation model output: a review of methods and limitations, *Prog. Phys. Geogr.*, *21*, 4, 530–548.
- Wilks, D. S. (1999), Multisite downscaling of daily precipitation with a stochastic weather generator, *Clim. Res.*, *11*, 125–136.
- Zorita, E., and H. von Storch (1999), The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods, *J. Climate*, *12*, 2474–2489.