



# The impact of soil moisture on precipitation downscaling in the Euro-Mediterranean area

Elke Hertig<sup>1</sup> · Yves Tramblay<sup>2</sup> · Karin Romberg<sup>1</sup> · Irena Kaspar-Ott<sup>1</sup> · Christian Merckenschlager<sup>1</sup>

Received: 15 September 2017 / Accepted: 8 June 2018 / Published online: 20 June 2018  
© The Author(s) 2018

## Abstract

In the present study the role of soil moisture (SM) in a statistical downscaling framework for precipitation in the Euro-Mediterranean domain is assessed. Different settings of the statistical downscaling models, differing in terms of the predictor variables used, are compared to quantify the influence of SM on the downscaling results. Results indicate an improvement of the skill of the statistical models when using SM information. This improvement is only moderate when averaged over the whole Euro-Mediterranean domain, but for specific regions the gain in performance is substantial. Regional projections of precipitation under the RCP4.5 and RCP8.5 scenario are considerably modified when SM is used as additional predictor in comparison with results based on atmospheric predictors alone.

**Keywords** Statistical downscaling · Soil moisture · Precipitation · Europe · Mediterranean area

## 1 Introduction

Analysis of climate variability and climate change generally relies on simulation models. State-of-art earth system models (ESMs) comprise models of atmospheric and oceanic circulation as well as models representing land surface processes. Despite the advances in numerical modelling and the increased resolution of such models, there is still a substantial need to downscale the ESM output to regional and local scales. Different types of regionalization techniques have been developed which comprise regional climate models (RCMs) as well as statistical downscaling approaches (Maraun et al. 2010). One approach in statistical downscaling involves transfer functions linking a set of large-scale atmospheric variables (predictors) to regional climate variables (predictands) during an observational period. The observed predictors–predictand relationships are subsequently used to assess future regional climate change. This is done by the substitution of the observational predictors by ESM predictors.

Commonly used predictor variables include pressure-related variables like sea level pressure and geopotential heights, as well as predictors describing thermo-dynamic properties of the atmosphere such as atmospheric humidity. Thus, the predictors–predictand relationships explain the joint short-term variations arising from the direct influence of the predictors on the predictand. Modifications of these relationships due to slowly varying boundary conditions are not determined, and variables like soil moisture (SM) are usually neglected in statistical downscaling equations. Compared to atmospheric variables, SM has a longer memory: after a rainfall event SM exhibit a sudden increase followed by a smooth recession driven by evapotranspiration and drainage (Brocca et al. 2014; Hagemann and Stacke 2015), and its spatial distribution is affected by meteorological forcing and soil properties (Wang et al. 2017). SM is also an important driver of climate variability and climate change (Seneviratne et al. 2006, 2010) and it should therefore not be disregarded in the scope of downscaling exercises.

To date, the investigation of land surface-climate interactions is mostly done by model studies. The Global Land–Atmosphere Coupling Experiment (GLACE, Koster et al. 2004) investigated land–atmosphere coupling of 12 Atmospheric General Circulation Models (AGCMs). Hot spots of land–atmosphere coupling were found to be located in transitional zones between dry and wet climates. Seneviratne et al. (2006), studying projected changes in interannual

✉ Elke Hertig  
elke.hertig@geo.uni-augsburg.de

<sup>1</sup> Institute of Geography, University of Augsburg, Alter Postweg 118, 86135 Augsburg, Germany

<sup>2</sup> HydroSciences Montpellier (IRD, CNRS, University of Montpellier), Montpellier, France

climate variability during summer using RCM and GCM (General Circulation Model) simulations, highlight the Mediterranean area as a hot spot of land–atmosphere coupling. Zhang et al. (2008), assessing the land–atmosphere coupling in boreal summer using GLDAS (Global Land Data Assimilation System) land surface model data, also identify southern Europe as a region with strong land–atmosphere coupling. The explanation for these results is directly related to the dependency of evapotranspiration on SM in the models that is very reminiscent of the classical conceptual framework: in dry regions, the models' evapotranspiration is strongly controlled by SM, but its absolute value and variations are too small to impact climate variability. In wet regions, evapotranspiration is large, but not controlled by SM, thus SM has little impact on evapotranspiration. Only in transitional regions between dry and wet climates, both conditions are met for strong SM–climate coupling: a strong dependency of evapotranspiration on SM and large mean evapotranspiration (Koster et al. 2004; Seneviratne et al. 2010). This is the case in semi-arid regions such as the Mediterranean area, where direct evaporation from the soil plays an important role on the surface energy balance (Taylor 2015).

According to Goessling and Reick (2011), three mechanisms can be distinguished via which evaporation affects precipitation: moisture recycling, local coupling and circulation. Moisture recycling is related to the influence evaporation exerts on precipitation via the atmosphere's moisture budget. Van der Ent et al. (2010) show with ERA-Interim data that continental moisture recycling plays a significant role in Europe. Thus, in western Europe the continental precipitation recycling ratio is already about 30% which indicates moisture transport with continental origin, from North America or eastern Europe, depending on the wind direction (Van der Ent et al. 2010). Local coupling comprises the influence of evaporation on precipitation via the atmosphere's thermal structure. Local coupling can be both, positive or negative. Schär et al. (1999) find in a study with a regional climate model that summertime European precipitation in a region between the wet Atlantic and dry Mediterranean climate depends heavily upon the SM content. As responsible mechanisms the buildup of shallow boundary layers due to wet soils and low Bowen ratios, and consequently high values of relative humidity that lower the level of free convection are discussed. In addition, the presence of a positive feedback of radiative origin with larger net radiative flux over moist soils and the presence of synoptic-scale forcing play a role. The third mechanism comprises links between surface evaporation and circulation. Shukla and Mintz (1982) and Goessling and Reick (2011) assess, using global GCM sensitivity experiments, that in northern summer dry continent conditions lead to a

significant intensification of the continental thermal lows. Therefore, the westerlies over Eurasia are weakened which in turn results in drier conditions in the western parts of the continent (Goessling and Reick 2011). In short, the three mechanisms can be summarized as effect 1 = moisture only, effect 2 = moisture + temperature, and effect 3 = moisture + temperature + circulation.

Under climate change conditions, Seneviratne et al. (2006) find that land–atmosphere interactions increase climate variability in Central and Eastern Europe. The authors argue that climatic regimes in Europe shift northwards in response to increasing greenhouse gas concentrations, creating a new transitional climate zone with strong land–atmosphere coupling. Orłowsky and Seneviratne (2012) point in a time slice comparison (2081–2100 minus 1980–1999) of IPCC AR4 simulations to the role of depleting SM as the link between drying and heating in the Mediterranean area. Similarly, Berg et al. (2016) have shown that global aridity is enhanced by the feedbacks of projected SM decrease, associated with the land surface's response to climate and CO<sub>2</sub> change. However, Taylor et al. (2012) emphasize that a positive feedback of SM on precipitation dominates in state-of-the-art ESMs, which is not evident from observations. This may contribute to excessive simulated droughts. Besides Orłowsky and Seneviratne (2012) highlight the large inter-model diversity of the representations of land–atmosphere coupling in the GCMs. This highlights that when SM is used within a downscaling context, GCM biases have to be kept in mind.

Comprehensively, the recognition of SM as a potentially important variable influencing climate variability and climate change leads to the need to explicitly consider this variable as predictor in the context of statistical downscaling. To date the role of slowly varying boundary conditions in the framework of statistical downscaling has not systematically been assessed. However, there exists the well-founded assumption that the consideration of land surface–atmosphere–precipitation interactions might lead to an improvement of regional assessments of future climate change in Europe and the Mediterranean. The goal of this study is to quantify the role of soil moisture in a statistical downscaling framework for precipitation in the Euro-Mediterranean domain.

The paper is organized as follows: Sect. 2 describes the precipitation, SM and atmospheric data sets used in this study. Section 3 gives the details on the methods used with respect to the regionalization of precipitation, preparation and analysis of the predictor data and of the downscaling modelling procedure. Section 4 presents the results, particularly focusing on the role of SM in the statistical downscaling models. Section 5 contains the discussion and conclusions.

## 2 Data

### 2.1 Precipitation

Precipitation data are taken from the daily 0.25°E-OBS dataset version 14 provided by the European Climate Assessment and Dataset (ECA&D, Haylock et al. 2008). A European-Mediterranean (EU-MED) domain is selected, covering 65°N–25°N and 12°W–50°E. Data in the time period 1950–2009 are selected and filtered for missing values for each month separately. Grid cells containing more than 6 months with more than 2 missing days per month are removed. The filtering procedure results in 18,974 grid boxes to be used for subsequent analyses. Monthly means are calculated from the daily data and the monthly data is grouped into 3 months seasons, each season shifted by 1 month (Jan/Feb/Mar, Feb/Mar/April, Mar/April/May, etc.).

### 2.2 Soil moisture data

Soil moisture data from the 3-hourly Global Land Data Assimilation System (GLDAS) dataset (Rodell et al. 2004) for the period 1950–2009 are used in this study. GLDAS provides near-real-time estimates of soil moisture fields derived from different uncoupled high-resolution land surface model (LSM) integrations incorporating satellite- and ground-based observations. We use the data from the 1.0° × 1.0° NOAH model integration, which provides four soil layers from 0 to 200 cm depth. For feedbacks between soil moisture and climate not only the surface SM, but the moisture content within the root zone (or the top meter) is relevant (Seneviratne et al. 2010). Thus, we integrate the data down to the depth of 100 cm using the trapezoidal method (e.g. Mittelbach et al. 2012). Also grid cells with soil frost for all days in a month are removed from the dataset. In the scope of using SM as a predictor for statistical downscaling, SM data are interpolated to a horizontal resolution of 2.0° to be consistent with the resolution of the current generation of ESMs and SM fields are cut to the EU-MED domain. This results in 177 grid cells available for further analyses. As for precipitation, monthly means are calculated and the months are grouped into 3-month seasons.

ESM data is taken from the MPI-ESM-MR model (Max Planck Institute Earth System Model running on medium resolution grid) and from the CNRM-CM5 model [developed jointly by CNRM-GAME (Centre National de Recherches Météorologiques—Groupe d'études de l'Atmosphère Météorologique and Centre Européen de Recherche et de Formation Avancée)]. The present study

should be seen as an exploratory study, and thus does not claim to include all available GCMs from CMIP5. The first ensemble member of the historical and the RCP4.5 and RCP8.5 scenario (Van Vuuren et al. 2011) runs performed for the Coupled Model Intercomparison Project round 5 (CMIP5) were downloaded from the CMIP5 archive. We use the period 1950–2005 of the historical runs and the period 2006–2100 of the scenario runs. The model output data is also interpolated to a horizontal resolution of 2.0°. MPI-ESM includes the JSBACH land surface scheme where SM is represented by a single-layer whose maximum depth is spatially varying. The maximum water depth corresponds to the root zone, and no water below is considered (Hagemann and Stacke 2015). Soil hydrology in CNRM-CM5 is represented by three soil layers in its land surface component, ISBA. The first layer is 1 cm deep. The two other layers have depth varying in space, depending on the vegetation types. The two layers distinguish between the rooting depth and the total soil depth (Voldoire et al. 2013). In the present work we use total soil moisture content. Only a few studies have validated the soil moisture produced by CMIP5 models. Yuan and Quiring (2017) compared near surface soil moisture from 17 CMIP5 models with in situ observations and ESA-CCI satellite soil moisture data. They observed good correlations with observations, except for a wet bias in the deeper soil layer during months when the soil is dry.

### 2.3 Atmospheric predictors

Geopotential height of the 700 hPa level as well as specific and relative humidity of the 850 hPa level for the time period 1950–2009 from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis (Kalnay et al. 1996) are considered as atmospheric predictors in the downscaling models. The domain for 700 hPa geopotential height is set to 70°N–20°N and 70°W–50°E in order to include the main centers of action important for precipitation in the EU-MED area. For the humidity variables, the smaller domain 65°N–27.5°N and 12.5°W–52.5°E is chosen. The horizontal resolution of the atmospheric variables is 2.5° × 2.5°, resulting in 1029 and 432 grid cells, respectively. The chosen variables already have been shown to be skillful predictors for precipitation in earlier studies (e.g. Hertig and Jacobeit 2008, 2013; Hertig et al. 2012).

The atmospheric variables are also extracted from the MPI-ESM-MR and CNRM-CM5 model runs. The model data is interpolated to the 2.5° resolution of the reanalysis data for the corresponding domains and monthly data is grouped to 3-month seasons.

### 3 Methods

#### 3.1 Regionalization of precipitation

S-mode principal component analysis (PCA, e.g. Preisendorfer 1988; von Storch and Zwiers 1999) with VARIMAX-rotation is applied to precipitation of each 3-month season in order to group grid boxes to regions with similar precipitation variability. Higher PC loadings can be used to specify the spatial location of the precipitation regions and the corresponding PC scores are used as the regions' predictand time series in the regression models. The selection of the number of principal components (PCs) is based on the criterion that each retained PC has to be representative for at least one input variable, following Jacobeit (1993) and Philipp et al. (2007). A PC is considered representative if the loading of this PC is larger than the loadings of the other PCs at a specific grid box by at least one standard deviation of all loadings at this grid box; furthermore, this loading has to be statistically significant at the 95% level. In addition, in order to prevent small regions, a further criterion is a minimum of five grid boxes for each precipitation region. The PC scores are subsequently used as predictand time series within the regression models (see Sect. 3.4).

#### 3.2 Principal component analyses of predictor data

In order to reduce dimensions of the data s-mode VARI-MAX-rotated PCA is also applied to the SM data sets as well as to the atmospheric predictor variables. PCA is applied to the predictor fields of each 3-month season separately. The resulting PC scores are used as predictor time series in the regression models. The PC loadings depict the spatial representation of the PCs and higher loadings show the spatial centers of variation. The number of PCs to be extracted follows the same procedure as described in Sect. 3.1 with the additional criterion of a minimum of 10 grid points for each predictor center of variation. The additional criterion is set in order to obtain predictors of skillful scale (Benestad et al. 2008) that are representable by the ESMs.

#### 3.3 Comparison of soil moisture data sets

Since the historical runs of the CNRM-CM5 and MPI-ESM-MR models are freely evolving climate simulations, there is no temporal correspondence with GLDAS data driven by reanalysis. Consequently, the comparison of the different soil moisture datasets is performed based on the distributions of seasonal means over 3 months. A measure of the discrepancy between two distributions can be obtained with quadratic statistics such as the Cramér–von Mises statistic

(Anderson and Darling 1952; Anderson 1962). The Cramér–von Mises (CM) statistic is a measure of the mean squared difference between cumulative distribution functions. Small values of the CM statistic indicate a small distance between the two distributions (Laio 2004). The CM statistic can be employed to measure the distance between two unspecified continuous distributions, as it is the case in the present work:

$$T = [NM/(N + M)] \int_{-\infty}^{\infty} [F_n(x) - G_m(x)]^2 dH_{n+m}(x) \quad (1)$$

where  $F_n$  and  $G_m$  are the empirical distributions of size  $n$  and  $m$ , and  $H_{n+m}(x)$  is the empirical distribution function of the two samples together. The null hypothesis, that  $F_n(x)$  and  $G_m(x)$  come from the same (unspecified) continuous distribution, is rejected when  $T$  exceeds a certain critical value. The critical values for  $T$  are given in Anderson (1962).

#### 3.4 Downscaling modeling procedure

All analyses are based on monthly data grouped into 3-month seasons. As downscaling technique, multiple linear regression analysis (MLR, for a detailed description see von Storch and Zwiers 1999) is used. The simultaneous use of SM and precipitation would probably reflect the impact of rainfall on SM rather than the other way around. Therefore, the SM feedback on precipitation can only be inferred when SM is leading. Thus, in the regression models SM is introduced with 1 up to 3 months lead-time. In order to exclude the possibility that monthly autocorrelation of precipitation will lead to an artificial model skill, the Ljung–Box test statistic (Ljung and Box 1978) with 1% significance level is used to assess whether the precipitation time series are independently distributed. The atmospheric variables (geopotential height, specific and relative humidity) are considered with no time lag, they give the instantaneous influence of the atmosphere on precipitation. As an example precipitation in Jan/Feb/Mar is used as predictand, the atmospheric variables in Jan/Feb/Mar and soil moisture in Dec/Jan/Feb, in Nov/Dec/Jan, and in Oct/Nov/Dec are used as predictors. Downscaling models are set up for each 3-month season, each season shifted by 1 month, thus yielding 12 seasonal analyses in total. In order to develop robust downscaling models the following procedure is applied:

- The subsequent steps are done using 1000 random samples with 2/3 of the data taken for calibration and the other 1/3 of the data taken for validation.
- At first, all potential predictor variables are used as input in the MLR. MLRs are run several times with one predictor variable left out in each case and the mean squared

error [MSE, see Eq. (4) below] in calibration and validation is calculated.

- All potential predictor variables are correlated with each other. It is assumed that the predictor left out with the highest MSE has the greatest influence on the predictand. Each couple of correlated predictors that exceed  $r = \text{abs}(0.4)$  is checked for the MSE from the last step. The predictor with the higher MSE remains as important one, the predictor with the lower MSE is removed from the predictors list. The remaining predictors are uncorrelated (below the defined threshold of 0.4).
- Stepwise MLR with the new set of uncorrelated predictors is performed. The AIC (Akaike information criterion, Akaike 1974) is used as selection criterion to add or remove terms. Predictors, which are significant in more than 1/3 of all the 1000 regression models, enter the final predictor set. Significance of a predictor in the stepwise MLR is based on the t-test of the regression coefficient with significance level of 0.1.
- MLR is re-run using the final predictor set and the following performance measures are calculated:

$$\text{Coefficient of determination, } R^2 = \frac{\sum (y^* - \text{mean}(y))^2}{\sum (y - \text{mean}(y))^2} \quad (2)$$

with  $y$ : observed precipitation value  
 $y^*$ : modelled precipitation value.

$$\text{Mean squared error skill score, } MSESS = \left( 1 - \frac{MSE_{\text{modelled}}}{MSE_{\text{reference}}} \right) \times 100 \quad (3)$$

with:

$$MSE_{\text{modelled}} = \frac{\sum (y^* - y)^2}{n} \quad (4)$$

$$MSE_{\text{reference}} = \frac{\sum (\text{mean}(y) - y)^2}{n} \quad (5)$$

$\text{mean}(y)$ : mean over the observations in the calibration/validation period

$n$ : number of observations.

$MSESS < 0$  indicates no skill, whereas  $MSESS = 100\%$  implies a perfect model.

Also the Pearson correlation coefficient is computed between modelled and observed precipitation time series. The final model performance is indicated by the mean over the 1000 random samples for each performance measure for calibration and validation.

Moreover, the downscaling procedure outlined above is compared to another model selection approach- regression models are also built using regression shrinkage and selection via the lasso (“least absolute shrinkage and selection

operator”, Tibshirani 1996; Tibshirani et al. 2012). Lasso tries to enhance prediction accuracy by shrinking or setting to 0 some coefficients and determines a subset of variables with the strongest effects. Within lasso a 10-fold cross validation is used to decide on the value of the tuning parameter  $t$ , with  $t$  controlling the amount of shrinkage that is applied to the estimates. As for the downscaling procedure outlined above, lasso regression is applied using 1000 random samples with 2/3 of the data taken for calibration and the other 1/3 of the data taken for validation.

Five different settings are used to investigate the influence of SM as predictor in precipitation downscaling models. The settings differ in terms of the predictors used:

1. Statistical models using all predictor variables (geopotential height, specific humidity, relative humidity, and SM).
2. Statistical models using only the circulation predictor (geopotential height). This setting is used to assess the circulation-based influences on precipitation.
3. Statistical models using all atmospheric predictors (geopotential height, specific and relative humidity). This setting is used to identify the atmospheric circulation and thermo-dynamic centers of variation which impact on precipitation.
4. Statistical models using geopotential height and SM. This setting is used to identify SM-circulation-precipitation relationships without considering atmospheric humidity.
5. Statistical models using only SM. This setting is used to assess the isolated downscaling skill of SM as predictor.

The predictors-precipitation relationships established in the observational period are subsequently used to assess the response of precipitation to changes of the predictors. For this purpose the ESM model data of the historical runs, the RCP4.5 and RCP8.5-scenario runs are projected in each case onto the existing predictor PCs of the observational period to obtain new predictor time series.

The assessed precipitation PC time series are subsequently back-transformed to the original E-OBS grid resolution and de-normalized to the original precipitation scale.

Projections regarding possible future precipitation changes under increased greenhouse gas conditions are performed and subsequently evaluated. The projections use the statistical relationships of the organized short-term atmospheric variability with precipitation and also show modifications of these relationships due to slowly varying boundary conditions.

## 4 Results

### 4.1 Precipitation regions and centers of variation of SM and atmospheric predictors

PCA of precipitation and predictor data have been applied to reduce dimensions of the data. Depending on the season, regionalization of the gridded E-OBS dataset by means of PCA yields 11 (winter) up to 24 (summer) regions with total amounts of explained variance from 61.7 to 70.7%. A summary of the number of PCs extracted and the amount of explained variance for all seasons can be found in Table 1. The PC scores are used as predictand time series in the MLR models.

**Table 1** Number of PCs and amount of explained variance from the PCAs of 3-month season precipitation

Season	Number of PCs	Amount of explained variance (%)
Jan/Feb/Mar	14	70.7
Feb/Mar/Apr	17	69.6
Mar/Apr/May	14	63.0
Apr/May/Jun	14	61.7
May/Jun/Jul	21	67.2
Jun/Jul/Aug	24	67.7
Jul/Aug/Sep	17	63.4
Aug/Sep/Oct	18	69.9
Sep/Oct/Nov	16	69.5
Oct/Nov/Dec	13	67.8
Nov/Dec/Jan	12	67.0
Dec/Jan/Feb	11	68.2

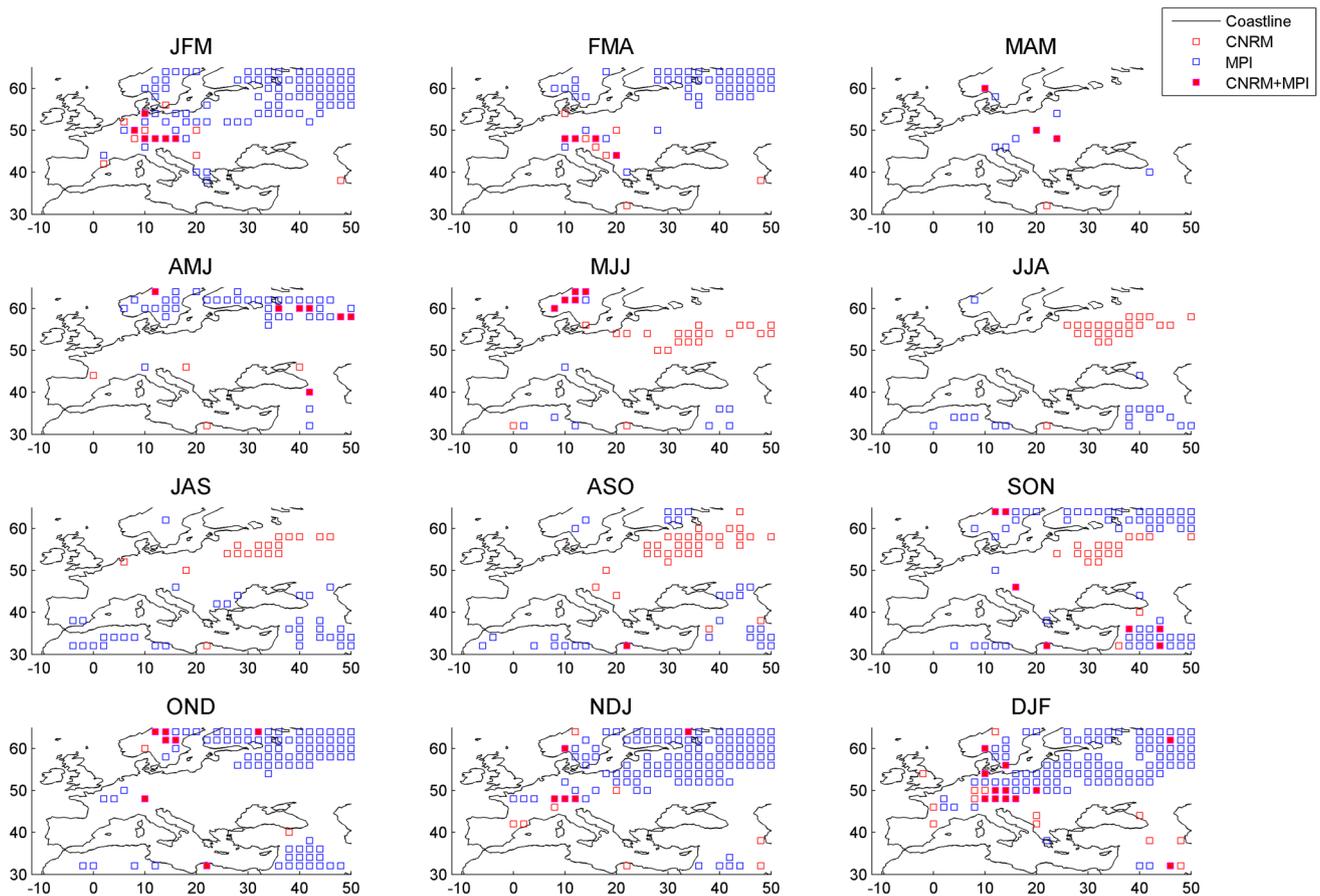
**Table 2** Number of PCs and amount of explained variance (Exvar) from the PCAs of 3-month season soil moisture (SM), 700 hPa geopotential height, 850 hPa specific humidity, and 850 hPa relative humidity

Season	SM		700 hPa geopotential height		850 hPa specific humidity		850 hPa relative humidity	
	Number of PCs	Exvar (%)	Number of PCs	Exvar (%)	Number of PCs	Exvar (%)	Number of PCs	Exvar (%)
Jan/Feb/Mar	8	64.4	9	90.0	7	74.7	6	66.5
Feb/Mar/Apr	8	70.2	9	90.1	6	76.8	10	69.6
Mar/Apr/May	8	77.5	8	90.0	6	80.3	10	65.0
Apr/May/Jun	7	77.8	8	90.6	6	80.9	8	64.3
May/Jun/Jul	6	74.6	8	88.5	7	76.6	9	69.5
Jun/Jul/Aug	7	73.6	11	82.0	11	65.6	13	60.0
Jul/Aug/Sep	8	69.4	8	85.3	10	70.1	10	67.3
Aug/Sep/Oct	6	66.3	6	84.5	6	73.6	9	72.2
Sep/Oct/Nov	5	69.0	6	85.2	9	84.5	9	71.3
Oct/Nov/Dec	5	69.6	7	88.4	8	84.8	12	65.3
Nov/Dec/Jan	8	72.8	9	90.6	7	79.7	7	68.0
Dec/Jan/Feb	8	67.0	11	86.1	7	75.8	6	66.7

Results of the PCAs of the predictor variables are given in Table 2. Seasonal PCAs of SM yield between 5 and 8 PCs with explained variances from 64.4 to 77.8%. Geopotential height of the 700 hPa level is condensed to 7 up to 11 PCs with explained variances of 82% up to 90.6%. From the PCAs of specific humidity 6–11 PCs with explained variances of 65.6% up to 84.8% are extracted, for relative humidity 6–13 PCs with explained variances from 60 to 72.2%. The PC scores are taken as predictor time series in subsequent MLR analyses.

### 4.2 Comparison of SM datasets

Figure 1 displays, for all 3-month seasons, the grid cells where the null hypothesis of the CM test is rejected at the 5% level, indicating that the distributions of SM simulated by CNRM-CM5 or MPI-ESM-MR are different from the GLDAS SM distribution. There are only a few cases where simulated SM from both CNRM-CM5 and MPI-ESM-MR models are not in agreement with GLDAS data. A maximum of 10 grid cells where the SM distributions are not the same is found during the Dec–Feb (DJF) season, in areas mostly located in central and north-eastern regions. For all seasons, there are a greater number of rejections of the null hypothesis with the MPI-ESM-MR model than with the CNRM-CM5 model, in particular between November and February for the north-eastern area of the domain covering the Baltic area and North Russia. Overall, it can be concluded that the three SM datasets are in good agreement for most parts of the EU-MED domain.



**Fig. 1** Grid cells where the SM distribution from GLDAS is significantly different from the SM distribution from CNRM-CM5 and/or MPI-ESM-MR according to the Cramér–von Mises test (5% significance level)

### 4.3 Performance of the statistical downscaling models

Five different predictor settings are used to investigate the potential of SM as predictor variable in the statistical downscaling models (see Sect. 3.4). Table 3 shows for the five different settings the model performance averaged across all precipitation regions for a season as well as the average across all regions and seasons (last column “Year” in Table 3). It can be seen that the best statistical model performance in calibration and validation (indicated as bold values in Table 3) is achieved for the predictor setting using all variables, including SM. Seasonal, over all precipitation regions averaged MSESS ranges between 34 and 57% ( $\bar{\varnothing}$  45%) in calibration and between 11 and 41% ( $\bar{\varnothing}$  28%) in validation for this predictor setting (Table 3). Only in four seasons (Apr–Jun, Jul–Sep, Aug–Oct, Sep–Nov) the MSESS in validation is higher when using only the atmospheric variables (geopotential height, specific and relative humidity) as predictors. The use of only atmospheric circulation (geopotential height) yields noticeably reduced MSESS,

pointing to the importance of atmospheric humidity and SM information in the models. In summary, the statistical downscaling models with all predictor variables (geopotential height, atmospheric humidity, and SM) result in the best performance in calibration as well as in validation. However, the gain in the model skill averaged over all precipitation regions in a season is only small when SM is included as predictor in addition to the atmospheric variables.

The use of the predictor setting geopotential height and SM (exclusion of humidity) yields a lower average model skill (see Table 3). This indicates that larger-scale atmospheric humidity carries information that is not included in the circulation and SM predictors. Thus, it is important to include atmospheric humidity in the precipitation models. Atmospheric humidity plays a role either by characterizing the atmospheric humidity content over the target region or via larger-scale humidity advection.

Looking at the isolated downscaling skill of SM as predictor shows  $R^2$  values of 0.09–0.18 and correlation coefficients between 0.18 and 0.37. This points to an influence of SM conditions on subsequent precipitation. However, the

**Table 3** Average performance over all precipitation regions per season for the five different predictor settings

	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ	DJF	Year
<b>hgt0700 + shum0850 + rhum0850 + SM-1 + SM-2 + SM-3</b>													
R <sup>2</sup> cal	<b>0.44</b>	<b>0.43</b>	<b>0.48</b>	<b>0.44</b>	<b>0.34</b>	<b>0.36</b>	<b>0.39</b>	<b>0.39</b>	<b>0.43</b>	<b>0.56</b>	<b>0.52</b>	<b>0.57</b>	<b>0.45</b>
MSESS cal	<b>44.10</b>	<b>42.95</b>	<b>47.78</b>	<b>43.81</b>	<b>34.17</b>	<b>35.92</b>	<b>38.63</b>	<b>39.10</b>	<b>42.58</b>	<b>56.15</b>	<b>52.24</b>	<b>56.78</b>	<b>44.52</b>
Correlation cal	<b>0.65</b>	<b>0.64</b>	<b>0.68</b>	<b>0.65</b>	<b>0.57</b>	<b>0.58</b>	<b>0.61</b>	<b>0.61</b>	<b>0.65</b>	<b>0.74</b>	<b>0.71</b>	<b>0.75</b>	<b>0.65</b>
R <sup>2</sup> val	<b>0.48</b>	<b>0.47</b>	<b>0.52</b>	<b>0.48</b>	<b>0.39</b>	<b>0.47</b>	<b>0.45</b>	<b>0.44</b>	<b>0.48</b>	<b>0.61</b>	<b>0.57</b>	<b>0.61</b>	<b>0.50</b>
MSESS val	<b>28.63</b>	<b>27.01</b>	<b>31.73</b>	28.68	<b>18.81</b>	<b>11.37</b>	20.57	22.05	25.38	<b>38.27</b>	<b>37.04</b>	<b>40.67</b>	<b>27.52</b>
Correlation val	<b>0.56</b>	<b>0.54</b>	<b>0.58</b>	0.56	<b>0.47</b>	<b>0.45</b>	<b>0.50</b>	0.50	0.54	<b>0.64</b>	<b>0.62</b>	<b>0.65</b>	<b>0.55</b>
<b>hgt0700</b>													
R <sup>2</sup> cal	0.32	0.27	0.31	0.28	0.20	0.22	0.25	0.22	0.23	0.32	0.36	0.41	0.28
MSESS cal	31.86	27.32	30.76	28.02	19.72	22.09	24.63	21.86	22.73	31.85	36.43	41.18	28.21
Correlation cal	0.54	0.49	0.53	0.51	0.41	0.45	0.47	0.44	0.45	0.53	0.57	0.62	0.50
R <sup>2</sup> val	0.35	0.31	0.34	0.32	0.24	0.28	0.29	0.25	0.26	0.36	0.40	0.45	0.32
MSESS val	21.73	17.85	21.09	18.01	10.35	5.96	12.73	12.38	15.86	23.44	26.77	30.57	18.05
Correlation val	0.47	0.42	0.48	0.44	0.34	0.34	0.38	0.37	0.42	0.48	0.51	0.56	0.43
<b>hgt0700 + shum0850 + rhum0850</b>													
R <sup>2</sup> cal	0.41	0.40	0.43	0.41	0.31	0.32	0.35	0.36	0.39	0.52	0.48	0.53	0.41
MSESS cal	40.71	40.35	43.25	41.21	30.79	31.95	35.12	35.67	39.18	52.21	47.80	53.13	40.95
Correlation cal	0.62	0.62	0.64	0.63	0.53	0.55	0.58	0.58	0.62	0.71	0.68	0.72	0.62
R <sup>2</sup> val	0.45	0.44	0.47	0.45	0.35	0.41	0.41	0.40	0.43	0.57	0.52	0.57	0.46
MSESS val	28.55	26.20	31.20	<b>29.12</b>	18.06	10.94	<b>20.96</b>	<b>24.03</b>	<b>25.98</b>	37.19	35.08	39.93	27.27
Correlation val	0.54	0.53	0.57	0.56	0.45	0.44	0.49	0.50	0.54	0.63	0.60	0.64	0.54
<b>hgt0700 + SM-1 + SM-2 + SM-3</b>													
R <sup>2</sup> cal	0.39	0.32	0.35	0.31	0.24	0.24	0.28	0.25	0.29	0.35	0.40	0.46	0.32
MSESS cal	39.26	32.17	35.19	30.77	23.61	24.50	28.00	24.94	29.10	34.76	39.72	45.54	32.30
Correlation cal	0.61	0.54	0.58	0.53	0.46	0.48	0.51	0.48	0.52	0.56	0.61	0.66	0.54
R <sup>2</sup> val	0.43	0.36	0.39	0.35	0.28	0.32	0.33	0.29	0.33	0.39	0.43	0.49	0.37
MSESS val	25.08	20.29	23.00	19.35	12.33	8.48	13.98	12.47	16.59	24.09	28.35	33.04	19.75
Correlation val	0.52	0.46	0.50	0.46	0.38	0.38	0.41	0.38	0.44	0.50	0.53	0.58	0.46
<b>SM-1 + SM-2 + SM-3</b>													
R <sup>2</sup> cal	0.13	0.11	0.14	0.14	0.09	0.09	0.11	0.14	0.15	0.12	0.10	0.10	0.12
MSESS cal	12.97	11.06	14.34	13.50	9.42	8.72	10.73	14.46	14.50	11.70	10.11	10.09	11.80
Correlation cal	0.35	0.31	0.36	0.36	0.29	0.28	0.31	0.37	0.36	0.33	0.30	0.30	0.33
R <sup>2</sup> val	0.16	0.14	0.18	0.17	0.13	0.15	0.15	0.18	0.18	0.15	0.14	0.13	0.16
MSESS val	2.74	2.14	1.22	0.55	0.19	-4.45	-0.87	-0.94	0.97	0.94	0.02	0.47	0.25
Correlation val	0.27	0.24	0.25	0.24	0.20	0.18	0.20	0.22	0.24	0.23	0.21	0.22	0.22

Bold values: best performance per season across the predictor settings

*hgt0700* geopotential height of the 700 hPa level, *shum0850* specific humidity of the 850 hPa level, *rhum0850* relative humidity of the 850 hPa level, *SM-1* soil moisture with 1 month lead time, *SM-2* soil moisture with 2 months lead time, *SM-3* soil moisture with 3 month lead time, *R<sup>2</sup>* coefficient of determination, *MSESS* mean squared error skill score, *cal* calibration, *val* validation

low MSESS values show that generally SM cannot be used as a prime predictor in downscaling models, but instead has to be seen as an additional predictor besides atmospheric circulation and humidity.

In summary the results show that there is a small benefit of the overall model performance when including SM as an additional predictor in the downscaling models. Yet one might argue that a small improvement of the overall model performance does not justify the extra effort in the

downscaling procedure. However, when looking at the individual results for each region and season several cases stand out where SM contributes to the downscaling model performance in a considerable way. Indeed, autocorrelation of precipitation might indirectly affect the downscaling model skill. But according to the Ljung–Box test statistic (1% significance level) applied to the precipitation time series of all 3-month seasons, monthly autocorrelation affects only 10 out of a total of 191 precipitation regions. Furthermore,

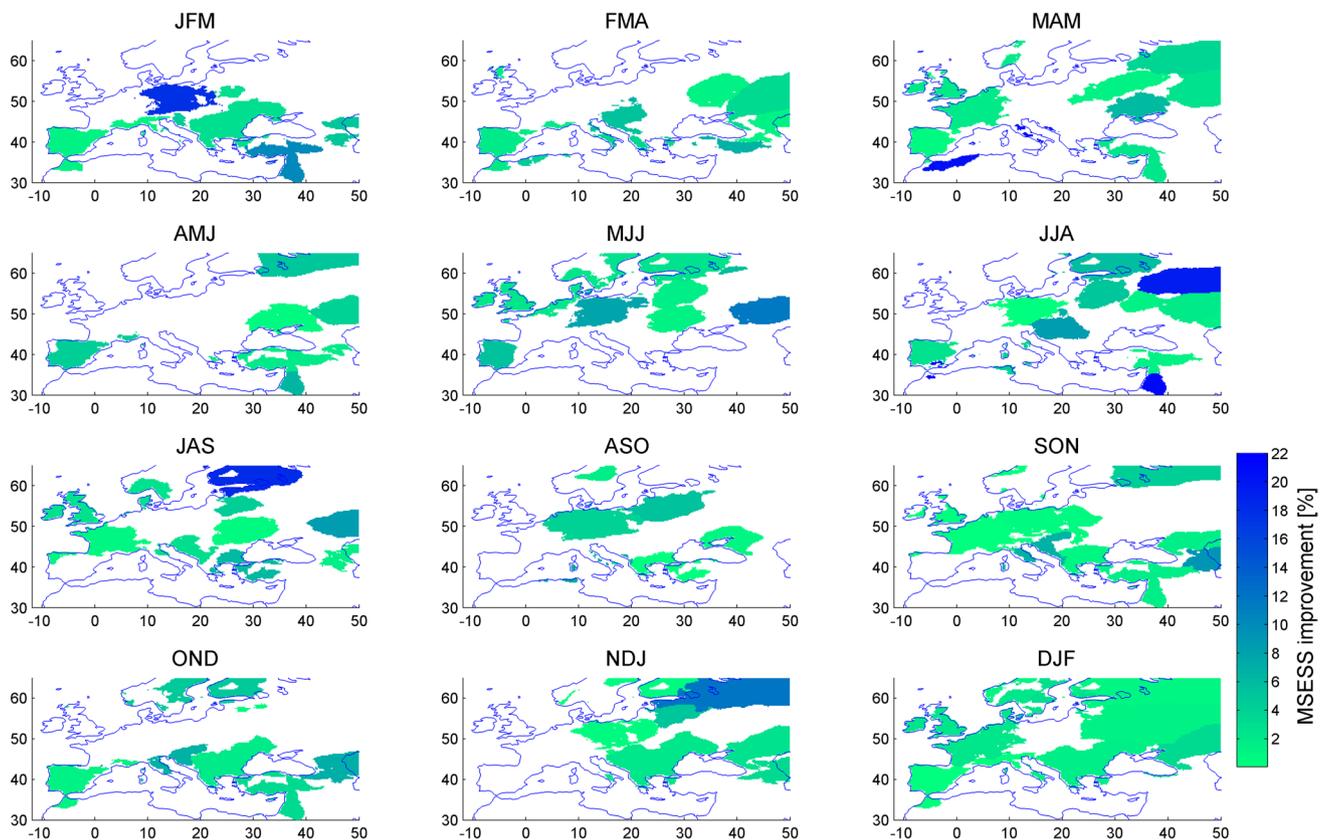
only two out of the ten regions correspond to regions with noticeably enhanced model skill through the inclusion of SM as additional predictor. Thus, autocorrelation of precipitation can be neglected within the interpretation of the results. As shown in Fig. 2, for some regions there is an increase up to +22% in the absolute values of the MSESS such as in Central Europe in JFM (Jan–Mar), or eastern Baltic and northern Russia in JJA (Jul–Aug) and JAS (Jul–Sep). Situations where SM plays an important role for precipitation downscaling can be found throughout the whole year depending on the region, there is no preference for a specific season (Fig. 2).

Comparing the statistical modeling procedure of the present study with lasso regression yields the result that the model performance is mostly somewhat reduced when using lasso regression. On average across all downscaling models  $R^2$  is 0.10/0.12 lower and MSESS 1.19%/4.21% lower in calibration and validation, respectively. Correlation coefficients are about the same in calibration and 0.09 lower in validation. Thus, lasso regression performance is lower especially in validation. This is due to the optimization of the models within model build-up in calibration. In contrast, regression models of the present study are optimized towards

both, calibration and validation. The rationale behind this is that the models have to be set up robustly in calibration. But with respect to future projections the downscaling models also need to be transferable to other time periods, this being assessed by their performance in validation. With respect to the role of SM in the lasso regression models, the average model performance in calibration is best for the predictor setting using all variables, including SM. In validation a slightly better overall performance can be found for the predictor setting using atmospheric variables alone. However, also in lasso regression regions reappear which show a markedly enhanced model skill in calibration and validation when SM is included as additional predictor.

#### 4.4 Relationships of SM with precipitation

The MLR results for the different precipitation regions in the different seasons show a complex pattern of relationships of precipitation with SM, the atmospheric circulation and humidity. In order to reveal the physical mechanisms on how SM can impact on precipitation, two examples are described in more detail.



**Fig. 2** Regions where the downscaling performance in validation is improved when using soil moisture information. For each region where there is an improvement with soil moisture, the colors indicate the absolute increase in the mean squared error skill score (MSESS in %)

#### (a) Precipitation over Central Europe in Jan–Mar.

For the Central European precipitation region in Jan–Mar a pronounced gain in model skill becomes evident when using SM as additional predictor (see JFM in Fig. 2). SM centers are selected in the regression models as predictors with lead times of 3 months (i.e. four SM centers in Oct–Dec) and 2 months (i.e. two SM centers in Nov–Jan). In Oct–Dec the SM centers are located over the target region itself as well as over the western and southern parts of Europe. In Nov–Jan the selected SM centers are located over eastern Europe and the eastern MED. It is interesting to note that SM over almost the whole EU-MED domain plays a role for precipitation in Central Europe in winter. Regression relationships of SM with precipitation are positive for all centers, meaning that enhanced SM leads to subsequent positive precipitation anomalies in Jan–Mar. An exception is the SM center in the eastern MED, showing a negative relationship with precipitation over Central Europe. This is probably due to the opposite connection of the eastern MED to the circulation over the North Atlantic area. Induced by the large-scale circulation in winter wet conditions occur over Europe, whereas dry conditions prevail over the eastern MED, and vice versa. Apart from that, SM in the different parts of the domain positively feedback on precipitation in Central Europe via the atmospheric circulation. The 700 hPa geopotential height centers of variation which are chosen in the regression models as predictors show the typical centers of action of EU-MED wintertime atmospheric circulation, i.e. the North Atlantic Oscillation resulting in south-westerly and westerly flow into the target region and the Russian High with a flow from easterly directions. Centers of variation for humidity, which are selected in the regression models, are located in the upstream regions of the advected air masses. In summary, positive SM anomalies in autumn/early winter across Europe can contribute to enhanced winter precipitation in Central Europe via connections with the atmospheric circulation and humidity.

#### (b) Precipitation over Russia in Jun–Aug.

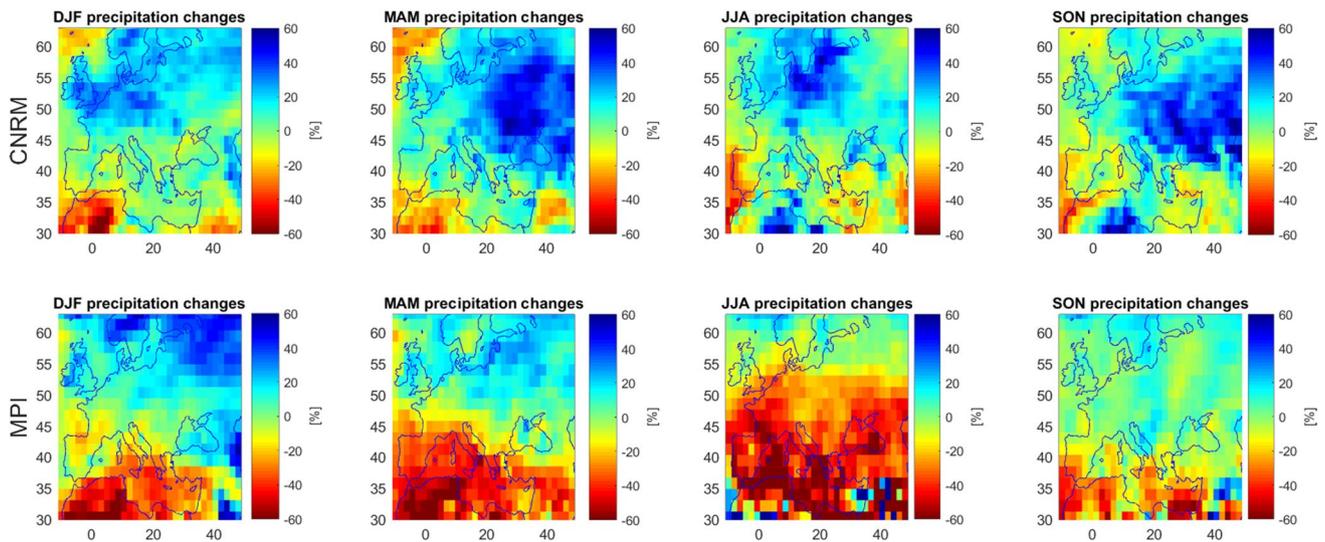
Figure 2 shows that in summer (Jun–Aug) the eastern and north-eastern areas of the domain contain regions where SM as predictor enhances MLR model skill. The SM-precipitation relationships are exemplarily described for the region with the highest gain in model skill (located over Russia, blue area in JJA in Fig. 2). In the MLR equation two SM centers of variation with 1 month lead time (SM in May–Jul) are selected. They are located over the Mediterranean area and over south-eastern and eastern Europe. The relationships with precipitation are positive, i.e. positive SM anomalies south of the target region lead to above normal precipitation amounts. The 700 hPa geopotential height and the 850 hPa relative humidity centers of variation which are dominant in the regression equation, are located over the target region itself. 850 hPa relative humidity variations

south of the target region are selected in the MLR equation as well. Thus, wetter than normal soils south of the target region can lead to an enhanced humidity advection from southerly directions by the correspondent atmospheric circulation anomalies, inducing positive rainfall anomalies over Russia in summer.

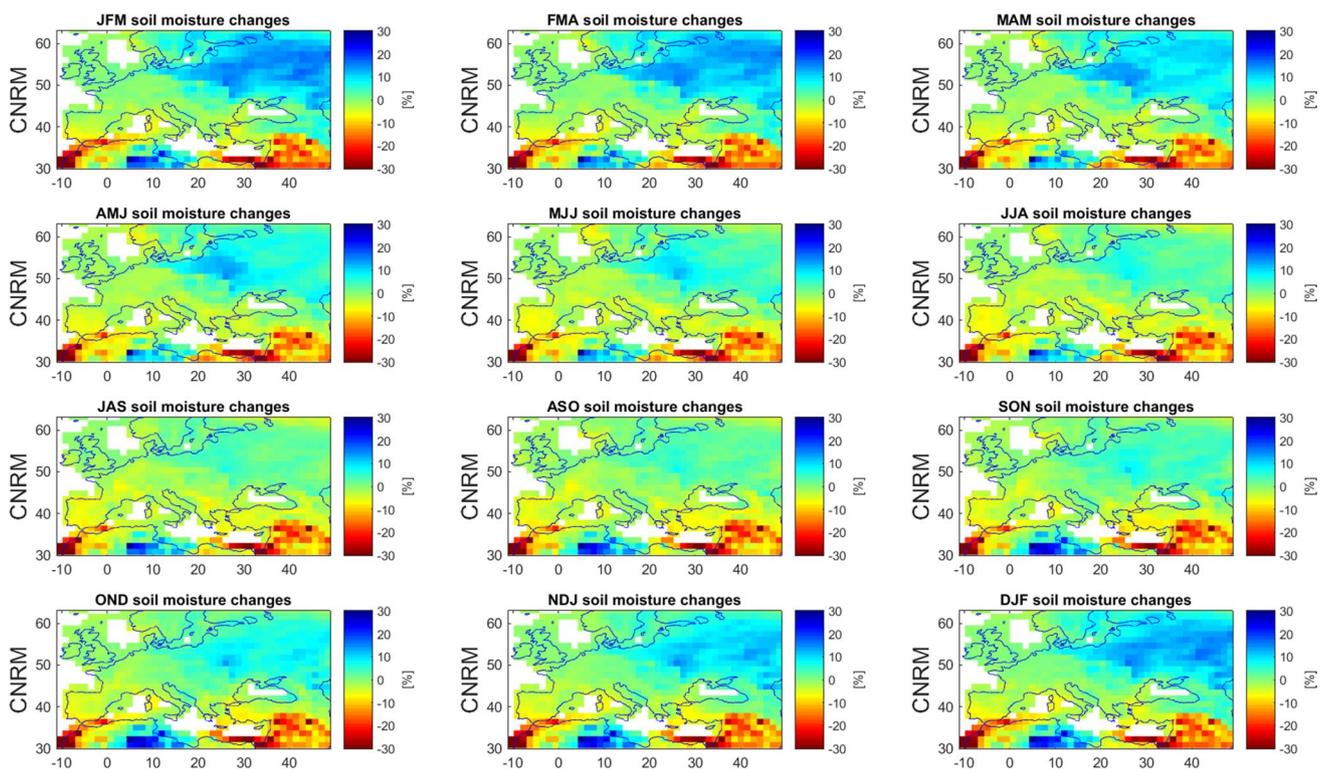
### 4.5 Change in precipitation when using SM as additional predictor

Although the scope of Perfect Prognosis statistical downscaling is to bypass the direct use of precipitation from the GCMs output, a comparison of the larger-scale precipitation change produced by the GCMs themselves with the signal given by the statistical downscaling methods may provide insight in the role of SM in the precipitation projections. Figure 3 shows the precipitation changes for the four seasons Dec–Feb (DJF), Mar–May (MAM), Jun–Aug (JJA), and Sep–Nov (SON) under the RCP8.5 scenario for the period 2071–2100 compared to the period 1971–2000 from the raw GCM output of CNRM-CM5 (Fig. 3, top) and MPI-ESM-MR (Fig. 3, bottom). Additionally, Figs. 4 and 5 display the SM changes in the two GCMs for all 3-month seasons. In CNRM-CM5 there are mainly precipitation increases in the extra-tropics, whereas decreases dominate in the MED. Stronger increases of SM are visible over north-eastern and eastern Europe in winter and spring and decreases over the eastern MED and western North Africa throughout the year. MPI-ESM-MR shows a somewhat different pattern of precipitation change, most notably a stronger drying over the southern parts of the domain particularly in MAM and JJA. SM is reduced over many parts of the MED area throughout the year, and over southern, central and eastern Europe in summer and autumn. In summary, in the direct GCM output in situ, contemporaneous SM and precipitation changes conform only to some extent, most likely because of the importance of other processes governing precipitation change. Also GCM deficiencies in the correct modeling of the complex SM-precipitation relationships have to be kept in mind.

The precipitation changes under the two scenarios RCP4.5 and RCP8.5 using the downscaling framework detailed above have been computed for the two ESMs, CNRM-CM5 and MPI-ESM-MR, for the periods 2036–2065 and 2071–2100 compared to the historical period 1971–2000. Results for the period 2071–2100 are displayed for the two ESMs on Figs. 6 and 7 for the RCP8.5 scenario in the four seasons. Three cases have been compared for the downscaling: statistical models with only the atmospheric circulation predictor (geopotential height, upper row in Figs. 6, 7), statistical models with all atmospheric predictors (geopotential height, specific and relative humidity, middle row in Figs. 6, 7), and statistical models with all predictor variables (including SM, bottom row in Figs. 6, 7). Note that



**Fig. 3** Precipitation changes under the RCP8.5 scenario for the period 2071–2100 compared to the period 1971–2000 from the direct GCM output of CNRM-CM5 (top) and MPI-ESM-MR (bottom)



**Fig. 4** Soil moisture changes under the RCP8.5 scenario for the period 2071–2100 compared to the period 1971–2000 from the direct GCM output of CNRM-CM5

SM is used with 1 up to 3 months lead-time in the statistical models. The patterns of change in the mid-twenty-first century (2036–2065) are very similar to the changes at the end of the century, but less intense. Thus, long-term precipitation

changes exhibit no major variability, but show a continuous progression during the course of the century. For the RCP4.5 scenario changes are less pronounced, but the spatial pattern of change is similar to the RCP8.5 scenario.

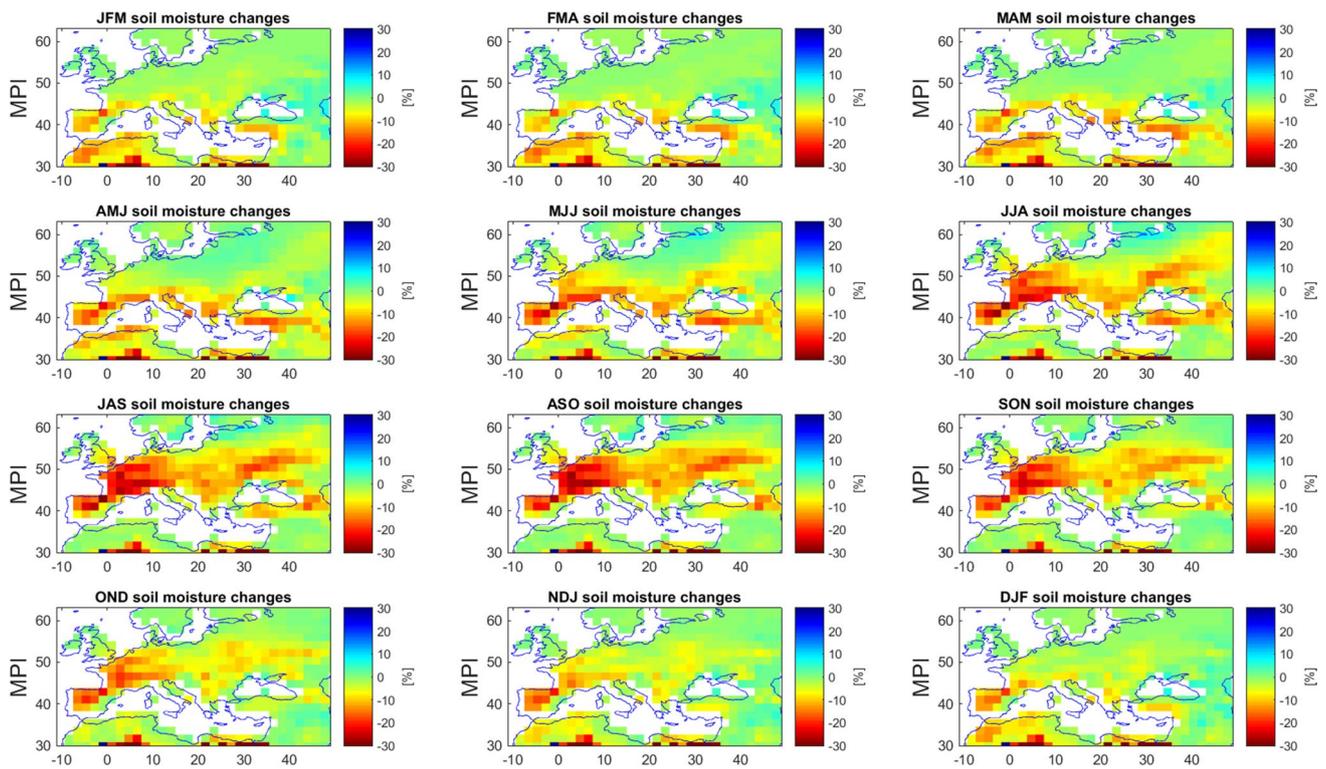


Fig. 5 Same as Fig. 4, but for MPI-ESM-MR

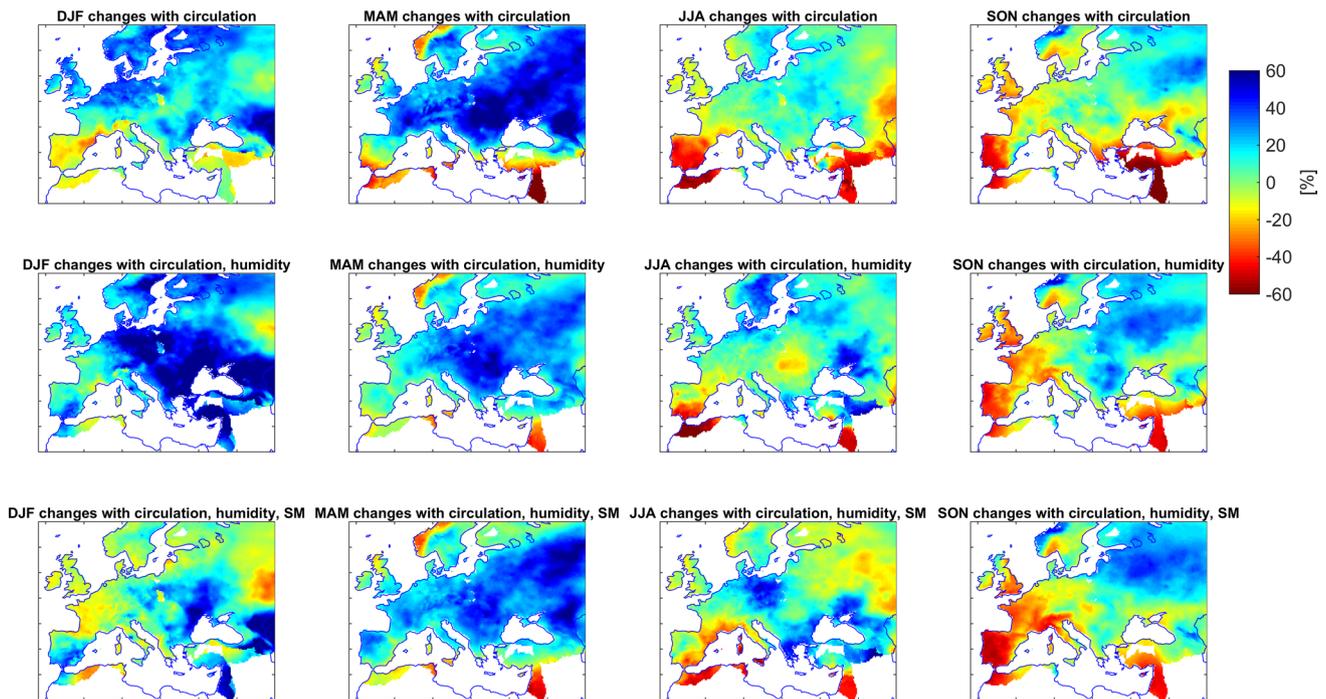
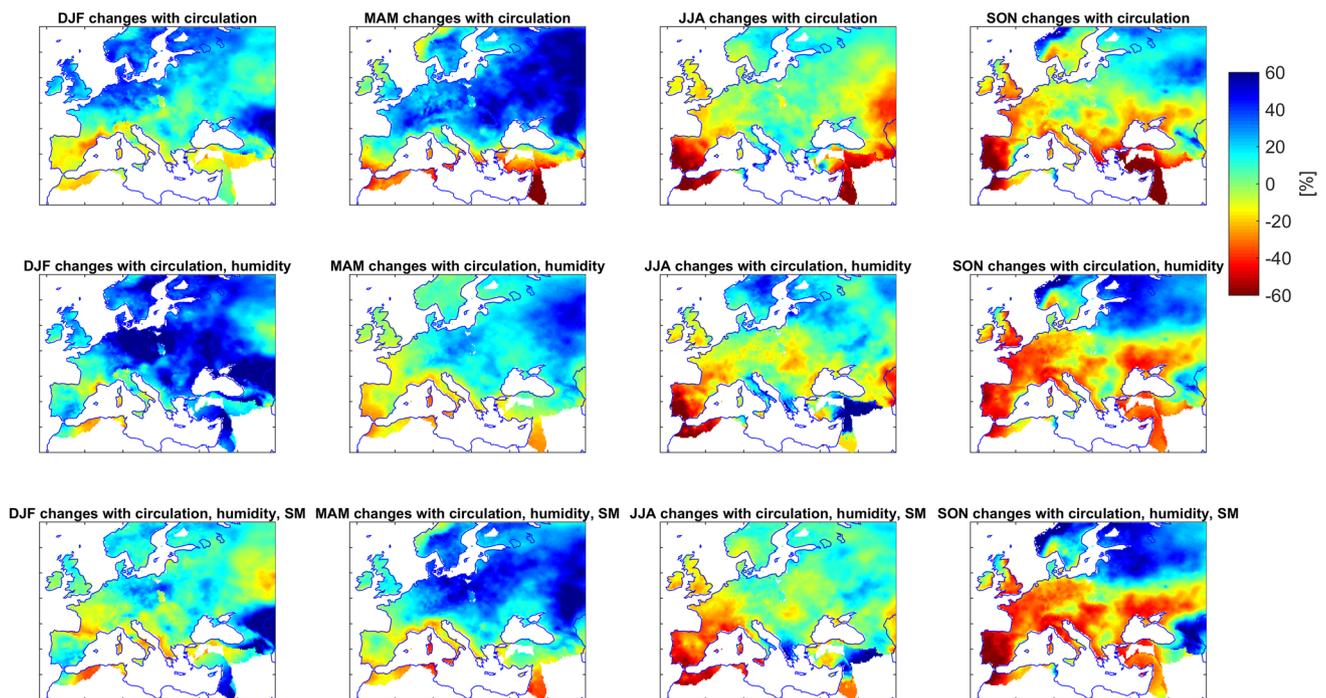


Fig. 6 Future changes in precipitation for 2071–2100 relative to 1971–2000, projected with the CNRM-CM5 model under the RCP 8.5 scenario. The upper row gives relative changes in precipitation using geopotential height as predictor in the downscaling model. The

middle row provides the changes using geopotential height, specific and relative humidity. The bottom row gives the changes using all predictor variables including SM



**Fig. 7** Same as Fig. 6, but for the MPI-ESM-MR model

Using only geopotential height as predictor, in DJF and MAM precipitation increases are visible over most parts of Europe, whereas decreases dominate over the MED. In JJA and SON, strong decreases occur over the western and eastern MED, together with increased precipitation over different parts of Europe, like central Europe in JJA and north-eastern Europe in SON. Using all atmospheric predictors (geopotential heights, specific and relative humidity), in DJF and MAM both models indicate a precipitation increase in central and eastern Europe. In DJF precipitation increases can be also seen over some southern regions, particularly over the eastern MED. This is due to the strong statistical model dependence of precipitation in this region with specific humidity over the northern and western parts of Europe. Humidity and geopotential height centers of variation often show a high correlation in the observational period and the question which variable center is included in the regression model is based on the statistical skill. Large changes of specific humidity in the GCM projections can then result in diverging downscaling results. Thus, even though the use of humidity results in a good statistical model quality in the observational period and this variable is regarded as an important driver of climate change, it may result in substantial modifications of the precipitation projections (see also Hertig and Jacobeit 2008).

In general, the magnitude of the changes is different from the two ESMs considered, however the spatial patterns of change generally agree. The consistency of the projected

precipitation patterns is due to the methodology on the one hand, since the same observational relationships between the large-scale predictors and regional precipitation are used for the projections. In addition, the projected precipitation patterns have to be explained from the predictor changes. For instance, in DJF MPI-ESM-MR and CNRM-CM5 show overall increases of specific humidity with largest values over northern Europe with values up to +40% and decreases of relative humidity over many parts of the MED with strongest values of -15% over the western MED, otherwise only small changes occur. Geopotential height exhibits increases, strongest over the western parts of the domain in MPI-ESM-MR and over north-eastern Europe in CNRM-CM5. Since the circulation (represented by geopotential height) and thermo-dynamic (represented by specific and relative humidity) changes are rather similar between the two GCMs, similar patterns of precipitation change are projected for the future.

When using SM in the downscaling approach (bottom row in Figs. 6, 7), the increase in precipitation over Europe in DJF is less pronounced. In this context, the precipitation projection from CNRM-CM5 predictors even leads to a decrease of precipitation over some parts of northern and western Europe. This can be explained by the observation-based precipitation-predictor links which contain specific lagged relationships of SM and precipitation. For northern Europe around the Baltic Sea there is a strong regression relationship of precipitation with SM in southern Morocco

and the eastern MED, with a lead time of SM of 2–3 months; [a geopotential height center of variation located over the British Isles and the North Sea as well as a specific humidity center of variation located over the Tyrrhenian Sea also play a major role, i.e. pressure variability west of the target region and atmospheric humidity from the northern Mediterranean act on precipitation over northern Europe. Besides, in situ SM in the target region is moderately strong connected with precipitation]. In the future assessments pronounced CNRM-CM5 SM reductions in western North Africa and the eastern MED in autumn and early winter (Fig. 4) induce precipitation decreases over northern Europe in winter. For western Europe precipitation is additionally positively connected with SM in the Iberian Peninsula and western Europe and negatively with SM in eastern Europe, i.e. above normal SM in the Iberian Peninsula and western Europe/eastern Europe in autumn and early winter are related to above/below normal precipitation over western Europe in winter; [in addition, two further geopotential height centers of variation, located over the eastern North Atlantic at about 30°N and over the eastern MED, as well as relative humidity over northern Africa and the eastern MED are selected in the regression equations]. In the projections MPI-ESM-MR shows strong SM reductions in the Iberian Peninsula and western Europe in autumn and early winter, CNRM-CM5 SM increases in eastern Europe, both inducing DJF precipitation decreases over western Europe. In summary, there are distinct changes in the future projections when SM is used as an additional predictor in the statistical downscaling models. In this regard the fundamental questions arise to what extent the SM-precipitation relationships are reliable and if these relationships can be transferred to future conditions using GCM output. These questions are taken up in the following section, again.

## 5 Discussion and conclusions

This study provides an assessment of the impact of soil moisture on precipitation downscaling in the Euro-Mediterranean area. The objective was to propose a statistical downscaling procedure to model precipitation considering SM as predictor in addition to commonly used atmospheric variables (atmospheric circulation represented by geopotential height of the 700 hPa level and atmospheric humidity typified by specific and relative humidity of the 850 hPa level). Five different settings of the downscaling models, differing in terms of the predictor variables used, have been compared to quantify the influence of SM on downscaled precipitation. Results indicate an improvement of the skill of the statistical models when using SM information. This improvement is only moderate when averaging

over the whole Euro-Mediterranean domain, but when looking at individual regions, the gain in performance can be substantial.

SM centers of variation selected in the regression models can match with the location of the target precipitation region, but can also be located in other parts of the domain. It points to the importance of regional coupling mechanisms as well as to the relevance of moisture advection via the atmospheric circulation. The relationships of SM with precipitation are mostly positive, i.e. wetter soils lead to enhanced precipitation amounts in the subsequent months. It should be highlighted however, that SM is used in this study as a large-scale predictor with a relatively coarse resolution, subgrid-scale local coupling mechanisms are not captured. The increase of statistical model performance when using SM as an additional predictor shows no dependence on the seasons and is rather scattered across the EU-MED domain. But it can be noted that in regions where there is a strong dependency of evaporation on SM and large mean evaporation values a stronger SM-precipitation coupling usually occurs. Nevertheless and importantly, also statistical models for regions and seasons which do not fit into the classical theoretical framework of SM-precipitation coupling can substantially benefit from the inclusion of SM as additional predictor. In these cases teleconnections between preceding SM anomalies and subsequent precipitation occur, with modifications of the large-scale atmospheric circulation and humidity playing an important role. In this regard the statistical modeling results diverge from the classical conceptual framework with local dependency of evaporation on SM, which is commonly used to understand SM-precipitation relationships in current GCMs (Seneviratne et al. 2010).

The statistical projections under climate change conditions are also impacted by the use of SM as an additional predictor. While there are mostly only small changes in spring, summer, and autumn, particularly winter climate is affected by the inclusion of SM. For instance winter increases in precipitation projected for most of the central, eastern and northern parts of Europe are substantially lower by comparison with the results obtained with a downscaling setting based only on atmospheric predictors. The statistically established SM-precipitation teleconnections considerably modify EU winter precipitation through SM signals from the southern parts of the domain. Under the use of CNRM-CM5 predictor output it even results in a contradictory pattern of change compared to the direct GCM precipitation output, the statistical projections using only atmospheric predictors, and also to other commonly known projection results of precipitation over Europe and the Mediterranean (e.g. CMIP5 models, Knutti and Sedláček 2013; EURO-CORDEX RCMs, Jacob et al. 2014).

Within these findings the following issues should be discussed and addressed for future research: (1) the statistical

relationships, established in the observational period, could be of limited suitability to be transferred to future conditions. Thus, the forcing of SM on precipitation could be overestimated. In combination with large changes of SM in the GCM projections it results in an unusual impact on the downscaling results. These issues should be investigated by detailed observation-based analyses on the possible local and remote effects of SM anomalies on atmospheric humidity, circulation patterns, and precipitation. In this regard teleconnection studies are needed which focus in particular on the physical processes of potential SM-precipitation links in the extended winter season. (2) There are shortcomings in the representation of SM forcing on precipitation in the GCMs. While in GCMs mostly the local land-atmosphere-climate interactions are focused on, the statistical results of this study suggest that also large-scale SM-atmosphere-precipitation teleconnections play an important role. One could argue that in the dynamical ESMs the influence of SM on precipitation should be completely included by modifications of the atmospheric circulation. Thus, atmospheric predictors alone would suffice to statistically downscale local/regional precipitation. If the relevant SM anomalies were already properly incorporated in the atmospheric link in ESMs, the differences in precipitation between the correspondent statistical model settings should be insignificant. However, we show in the present study that the explicit inclusion of SM in the downscaling equations impacts on the downscaling results. Already Rowntree and Bolton (1983) showed in a GCM study that a mid-latitude SM anomaly can have important effects on humidity and precipitation, not only over the anomaly but also through advection of modified air from the anomaly area to other regions. Since then only few studies have investigated the impacts of SM on air properties like atmospheric humidity and extratropical circulation patterns (Seneviratne et al. 2010). Thus, further modeling studies are needed to investigate the role of SM-atmosphere-precipitation teleconnections in dynamical models.

**Acknowledgements** Funding is provided by the DFG (German Research Foundation) under contract HE 6186/4-1. Further financial support was provided by the Bayerisch-Französisches Hochschulzentrum/Centre de Coopération Universitaire Franco-Bavarois.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
- Anderson TW (1962) On the distribution of the two sample Cramér–von Mises criterion. *Ann Math Stat* 33:1148–1159
- Anderson TW, Darling DA (1952) Asymptotic theory of certain ‘goodness of fit’ criteria based on stochastic processes. *Ann Math Stat* 23(2):193–212
- Benestad RE, Chen D, Hanssen-Bauer I (2008) Empirical-statistical downscaling. World Scientific Publishing Company, New Jersey
- Berg AM, Findell K, Lintner B, Giannini A, Seneviratne S, van den Hurk B, Lorenz R, Pitman A, Hagemann S, Meier A, Cheruy F, Ducharne A, Malyshev S, Milly PCD (2016) Land–atmosphere feedbacks amplify aridity increase over land under global warming. *Nat Clim Change* 6:869–874
- Brocca L, Ciabatta L, Massari C, Moramarco T, Hahn S, Hasenauer S, Kidd R, Dorigo W, Wagner W, Levizzani V (2014) Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *J Geophys Res Atmos* 119:5128–5141
- Goessling HF, Reick CH (2011) What do moisture recycling estimates tell us? Exploring the extreme case of non-evaporating continents. *Hydrol Earth Syst Sci* 15:3217–3235
- Hagemann S, Stacke T (2015) Impact of the soil hydrology scheme on simulated soil moisture memory. *Clim Dyn* 44:1731–1750
- Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high resolution gridded data set of surface temperature and precipitation for 1950–2006. *J Geophys Res* 113:D20119. <https://doi.org/10.1029/2008JD010201>
- Hertig E, Jacobeit J (2008) Assessments of Mediterranean precipitation changes for the 21st century using statistical downscaling techniques. *Int J Climatol* 28:1025–1045
- Hertig E, Jacobeit J (2013) A novel approach to statistical downscaling considering non-stationarities: application to daily precipitation in the Mediterranean area. *J Geophys Res Atmos* 118:520–533
- Hertig E, Seubert S, Paxian A, Vogt G, Paeth H, Jacobeit J (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* 111:1–20
- Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer LM et al (2014) EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg Environ Change* 14(2):563–578
- Jacobeit J (1993) Regionale Unterschiede im atmosphärischen Zirkulationsgeschehen bei globalen Klimaveränderungen. *Erde* 124:63–77
- 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 NCAR/NCEP 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–470
- Knutti R, Sedláček J (2013) Robustness and uncertainties in the new CMIP5 climate model projections. *Nat Clim Change* 3(4):369
- Koster R, Dirmeyer P, Guo Z, Bonan G, Chan E, Cox P, Gordon CT, Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu Ch H, Malyshev S, McAvaney B, Mitchell K, Mocko D, Oki T, Oleson K, Pitman A, Sud YC, Taylor Ch, Verseghy D, Vasic R, Xue Y, Yamada T (2004) Regions of strong coupling between soil moisture and precipitation. *Science* 305:1138–1140
- Laio F (2004) Cramér–von Mises and Anderson–Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resour Res*. <https://doi.org/10.1029/2004WR003204>
- Ljung GM, Box GEP (1978) On a measure of lack of fit in time series models. *Biometrika* 65(2):297–303

- Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* 48:RG3003. <https://doi.org/10.1029/2009RG000314>
- Mittelbach H, Lehner I, Seneviratne SI (2012) Comparison of four soil moisture sensor types under field conditions in Switzerland. *J Hydrol* 430–431:39–49
- Orlowsky B, Seneviratne SI (2012) Global changes in extreme events: regional and seasonal dimension. *Clim Change* 110:669–696
- Philipp A, Della-Marta PM, Jacobeit J, Fereday DR, Jones PD, Moberg A, Wanner H (2007) Long-term variability of daily north Atlantic–European pressure patterns since 1850 classified by simulated annealing clustering. *J Clim* 20:4065–4095
- Preisendorfer RW (1988) Principle component analysis in meteorology and oceanography. Elsevier, Amsterdam
- Rodell M, Houser PR, Jambor U, Gottschalck J, Mitchell K, Meng C-J, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin JK, Walker JP, Lohmann D, Toll D (2004) The global land data assimilation system. *Bull Am Meteorol Soc* 85:381–394
- Rowntree PR, Bolton JA (1983) Simulation of the atmospheric response to soil moisture anomalies over Europe. *Q J R Meteorol Soc* 109(461):501–526
- Schär Ch, Lüthi D, Beyerle U (1999) The soil-precipitation feedback: a process study with a regional climate model. *J Clim* 12:722–741
- Seneviratne S, Lüthi D, Litschi M, Schär C (2006) Land–atmosphere coupling and climate change in Europe. *Nature* 443:205–209
- Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, Orlowsky B, Teuling AJ (2010) Investigating soil moisture–climate interactions in a changing climate: a review. *Earth Sci Rev* 99:125–161
- Shukla J, Mintz Y (1982) Influence of land-surface evapotranspiration on the Earth's climate. *Science* 215:1498–1501
- Taylor CM (2015) Detecting soil moisture impacts on convective initiation in Europe. *Geophys Res Lett* 42:4631–4638
- Taylor CM, de Jeu RAM, Guichard F, Harris PP, Doriga WA (2012) Afternoon rain more likely over drier soils. *Nature* 489(7416):423
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B (Methodol)* 58:267–288
- Tibshirani R, Bien J, Friedman J, Hastie T, Simon N, Taylor J, Tibshirani RJ (2012) Strong rules for discarding predictors in lasso-type problems. *J R Stat Soc B (Stat Methodol)* 74(2):245–266
- Van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque JF, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK (2011) The representative concentration pathways: an overview. *Clim Chang* 109:5–31
- Van der Ent RJ, Savenije HHG, Schaefli B, Steele-Dunne SC (2010) Origin and fate of atmospheric moisture over continents. *Water Resour Res* 46:W09525
- Voltaire A, Sanchez-Gomez E, Méliá DS, Decharme B, Cassou C, Sénési S, Déqué M (2013) The CNRM-CM5.1 global climate model: description and basic evaluation. *Clim Dyn* 40(9–10):2091–2121
- von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge
- Wang T, Franz TE, Li R, You J, Shulski MD, Ray C (2017) Evaluating climate and soil effects on regional soil moisture spatial variability using EOFs. *Water Resour Res* 53:4022–4035. <https://doi.org/10.1002/2017WR020642>
- Yuan S, Quiring SM (2017) Evaluation of soil moisture in CMIP5 simulations over the contiguous United States using in situ and satellite observations. *Hydrol Earth Syst Sci* 21:2203–2218
- Zhang J, Wang WC, Wie J (2008) Assessing land–atmosphere coupling using soil moisture from the Global Land Data Assimilation System and observational precipitation. *J Geophys Res* 113:D17119