



Bias adjustment for decadal predictions of precipitation in Europe from CCLM

Jingmin Li^{1,3} · Felix Pollinger¹ · Hans-Juergen Panitz² · Hendrik Feldmann² · Heiko Paeth¹

Received: 5 July 2018 / Accepted: 23 January 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

A cross-validated model output statistics (MOS) approach is applied to precipitation data from the high-resolution regional climate model CCLM for Europe. The aim is to remove systematic errors of simulated precipitation in decadal climate predictions. We developed a two-step bias-adjustment approach. In step one, we estimate model errors based on a long-term ‘CCLM assimilation run’ (regionalizing data from a global assimilation run) and observational data. In step two, the resulting transfer function is applied to the complete set of decadal hindcast simulations (285 individual runs). In contrast to lead-time-dependent bias-adjustment approaches, this one is designed for variables with poor decadal prediction skill and without dominant lead-time-dependent bias. In terms of the CCLM assimilation run, MOS is shown to be effective in predictor selection, model skill improvement, and model bias reduction. Yet, the positive effect of MOS correction is accompanied with an underestimation of precipitation variability. After MOS application, an estimated mean square skill score of more than 0.5 is observed regionally. Simulated precipitation in decadal hindcasts is further improved when the MOS is trained on the basis of other decadal hindcasts from the same regional climate model but with a large underestimation in forecast uncertainty. Our results suggest that the MOS system derived from the assimilation run is less effective but allows the potential climate predictability in decadal hindcasts and forecasts to be retained. Using hindcasts itself for training is recommended unless a statistical method is capable of distinguishing biases and predictions within a hindcasts dataset.

Keywords Bias-adjustment · CCLM · Hindcasts · Decadal prediction · Precipitation · Model output statistics

1 Introduction

Scientific interest in decadal climate prediction, with a focus on the next 1–10 years, has significantly increased over the last years. The Coupled Model Intercomparison Project phase 5 (CMIP5) and the comprehensive project for mid-term climate forecast (MiKlip) are two major projects in

this field. The CMIP5 includes a set of decadal climate prediction experiments using multi-model ensembles (Meehl et al. 2014). The MiKlip founded by the German Federal Ministry for Education and Research, focuses on decadal climate prediction (Marotzke et al. 2016). Decadal climate predictions aim to provide important and crucial inputs for near-term policy and decision makers regarding e.g. water supply, yield estimation, energy resources (e.g. WMO 2011) and for other end users such as planners or hydrological modelers.

The decadal climate prediction community devotes various activities to improve model’s prediction skill, including the improvement of model setup and initial conditions (e.g. Kruschke et al. 2015; Pattatagus-Abraham et al. 2016; Paeth et al. 2018; Boer et al. 2016), by means of dynamical or statistical downscaling (e.g. Mieruch et al. 2014; Reijers et al. 2015; Paxian et al. 2016; Paeth et al. 2017), and performing bias-correction/recalibration post-processing (e.g. Kharin et al. 2012; Fuckar et al. 2014; Kruschke et al. 2015; Pasternack et al. 2018; Boer et al. 2016). For decadal

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00382-019-04646-y>) contains supplementary material, which is available to authorized users.

✉ Jingmin Li
jingmin.li@uni-wuerzburg.de

¹ Institute of Geography and Geology, University of Wuerzburg, Wuerzburg, Germany

² Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany

³ Institute of Atmospheric Physics, German Aerospace Center, Oberpfaffenhofen, Germany

prediction skill evaluation, Boer et al. (2016) recommend to use consistent time periods for the analysis of different lead years, to avoid uncertainty in the estimated skills due to different sample sizes and difficulties in forecast comparison among projects.

Among these activities, statistical bias-adjustment is the most cost-effective way. It aims to improve the model's prediction skill by reducing its systematic errors through post-processing. Systematic errors of decadal predictions may contain a lead time-dependent bias which is caused by imposing an initialized observation-based state on an unconstrained model (e.g. Meehl et al. 2014). To address this problem, a specific drift (lead-time-dependent bias) adjustment at yearly timescale has been developed and improved over the last years (Kharin et al. 2012; Fuckar et al. 2014; Kruschke et al. 2015; Pasternack et al. 2018). Although a part of the prediction uncertainty is inherent to the nature of the Earth's climate system (Hawkins et al. 2016), successful bias-adjustments have been used for decadal predictions of near-surface temperature (Kharin et al. 2012; Pasternack et al. 2018), large-scale sea surface temperature indices, the Northern Hemisphere sea ice extent (Fuckar et al. 2014), and Northern Hemisphere winter storms (Kruschke et al. 2015). So far, meaningful decadal prediction skills and bias-adjustment are limited to temperature relevant variables. The decadal prediction skill scores for precipitation are generally poor (e.g. Kadow et al. 2015; Boer et al. 2016). Also the application of bias-adjustment for decadal prediction of precipitation is missing. This study might be a first take on this topic.

Difference from the bias-adjustment methods summarized above, our approach is not lead-time-dependent. Instead, we estimate and adjust specific error components in decadal predictions such as the systematic model error or the errors in the predicted large scale pattern, based on what our statistical method learns from the training simulations. Our results can be used as a supplement and reference to the widely-used lead-time-dependent bias-adjustment methods for decadal predictions. In this study, we show three novelties in bias-adjustment of decadal predictions as follows. First, we address the bias-adjustment of systematic errors in decadal predictions with a different philosophy: we develop a two-step bias-adjustment approach. In the first step, we use statistical analysis to estimate model biases based on a downscaled long-term assimilation run, leaving the hindcasts themselves untouched. In the second step, we apply the estimated transfer function (*TF*) to all decadal hindcasts (285 in total) to adjust model biases individually. Second, our approach extends the 1D standard bias-adjustment to multi-dimensional bias adjustment. More specifically, bias-adjustment for precipitation is not only based on model simulated precipitation, but also on temperature, sea level pressure and their leading large-scale EOF patterns. Piani and Haerter (2012) and Cannon (2016) demonstrated that multivariate

bias-correction improves the 1D method by keeping the variables' dynamic dependence. Third, our approach extents previous decadal bias-adjustment methods at yearly timescale to monthly timescale. Recent publications have addressed the importance of advancing seasonal to inter-annual prediction (e.g., Buizza and Richardson 2017), and great progress has been achieved in recent decades (e.g. Kirshnamurti et al. 1999; Hudson et al. 2011; Smith et al. 2015). Although seasonal skill of decadal prediction is limited to the first few forecast years (e.g. Stockdale et al. 2011; Gonzalez and Goddard 2016), we still evaluate the month-to-month variation for a 10-year period. This evaluation might have no practical impact for the end user, but it gives insights into the sources of model errors and functionality of our bias-adjustment at monthly timescale.

This paper is structured as follows: Sect. 2 describes data and simulations. Section 3 explains the developed post-processing approach and our evaluation method. Section 4 demonstrates the efficiency of our MOS approach in terms of the downscaled long-term assimilation run regarding the following aspects: predictors selection, a gain of skill and an accompanied limitation. In Sect. 5, the *TF* is applied to the decadal predictions. The bias-adjustment is evaluated at two different timescales, inter- and intra-annual. The results are discussed in Sect. 6, followed by conclusions in Sect. 7.

2 Data and simulations

For the MOS approach we use data from the non-hydrostatic regional climate model CCLM (Rockel et al. 2008). The CCLM dynamically downscale data from the global MiKlip decadal prediction system, which uses the coupled Earth system model MPI-ESM-HR (Müller et al. 2018; Giorgetta et al. 2013; Jungclaus et al. 2013). The dynamical downscaling of the global model output is carried out at a horizontal resolution of 0.22° (~25 km) over Europe. The dynamical downscaling with CCLM has been demonstrated to provide a statistically significant added value in terms of model bias and decadal climate prediction skill compared with the driving global model (e.g. Panitz et al. 2014; Paxian et al. 2016; Paeth et al. 2017). Two types of CCLM simulations are considered in this study: a long-term CCLM simulation driven by a global MPI-ESM-HR assimilation run (for brevity the expression, we name it 'CCLM assimilation run' afterwards), and a large number of decadal hindcasts generated in ensemble mode.

2.1 CCLM assimilation run

Assimilation is a process of incorporating observations into a climate model. The model is assimilated in each internal simulation time step and nudged towards observations.

Hence, sources of model errors from initial conditions or forecast system are kept small and the occurring errors in the assimilation run can be assumed as the model's systematic bias. Therefore, a MOS system can be ideally trained with this assimilation run and the resulting *TF* can be transferred to other simulations with the same model, e.g. decadal predictions. The advantage is that this bias adjustment is independent of the potential forecast error due to the unpredictability of climate at given time scale.

The so called 'CCLM assimilation' run is realized by downscaling the global MPI-ESM-HR assimilation run for the period 1961–2000. The CCLM simulation does not include an assimilation by itself. The global model uses Ocean Reanalysis (ORAS4) (Balmaseda et al. 2015) data for the assimilation of the ocean temperature and salinity, and ERA40/ERA-Interim (Uppala et al. 2005; Dee et al. 2011) for the assimilation of the atmosphere, both datasets are provided by the European Center for Medium-range Weather Forecast (ECMWF).

2.2 CCLM decadal hindcasts

A hindcast is known as a re-forecast with climate models for a period in the past, and is a way of testing and evaluating a climate models' prediction skill. The model is integrated during past and progresses forward in time just like a forecast. The hindcasts are given as ensembles with slightly varying atmospheric initial conditions, aiming to obtain an adequate range of uncertainty for the predicted climate. CCLM decadal hindcasts represent our ultimate target of bias-adjustments. We use data from 285 (5 members \times 57 initializations) individual decadal hindcasts in total. The decadal hindcasts are initialized yearly for 1960–2016 and then integrated over 10 years, each with five ensemble members by downscaling five different realizations of MPI-ESM-HR hindcasts. They are initialized by MPI-ESM with anomaly-field forcing in the ocean (Müller et al. 2012; Pohlmann et al. 2013). In contrast to the drift problem of initial full-field forcing in decadal prediction (e.g. Sansom et al. 2016; Paeth et al. 2018), drifts in the anomaly-field cancels on average but might have opposite signs for early lead years (Kruschke et al. 2015).

Hindcast simulations are assumed to contain more sources of error than the assimilation run. Decadal hindcasts are initialized at one observational state but evolve freely over the next 10 years. They have forecast uncertainty and contain also a different type of systematic error: the lead-time dependent drift error (Meehl et al. 2014). To estimate the model deficiencies, model predictions for the variables precipitation and near-surface temperature, which typically are most relevant for end users are compared with the version 14.0 E-OBS gridded observations (Haylock et al. 2008). E-OBS data is provided by the European Climate

Assessment and Dataset project. Daily data is available from 1950 to 2017 for the region 25°N – 75°N \times 40°W – 75°E . Figure 1 shows 10-year long-term means of annual precipitation and temperature for the years 2001–2010 (five members of decadal hindcasts which are initialized in the year 2000 including lead year 1–10) over Europe from the RCM (CCLM), observational data (E-OBS) and their differences. This decade has been chosen because it is referred to as a retrospective forecast during which no observational data is applied for model calibration. CCLM basically reproduces the observed pattern of annual precipitation and temperature, but also shows some systematic bias with respect to E-OBS. CCLM overestimates the precipitation amount over the northern and eastern parts of Europe and underestimates precipitation in parts of southeastern Europe. There is a bias of more than 200 mm/year for Scandinavia and the eastern part of the model domain. These biases correspond to about 10–50% of present-day rainfall totals (Fig. 1b, c). In contrast to precipitation, CCLM exhibits a small bias in temperature prediction. Except for some southeastern sub-regions, the CCLM temperature bias from E-OBS is within ± 2 K (Fig. 1f). The systematic bias of precipitation is more critical than temperature. Besides, previous studies claimed that it is also more challenging to have a skillful decadal prediction for precipitation (e.g. Meehl et al. 2014; Kadow et al. 2015; Boer et al. 2016). Therefore, we focus on the bias-adjustment for precipitation.

3 Methods

3.1 Model output statistics (MOS)

Our MOS approach is based on a cross-validated step-wise multiple linear regression method. It performs a reasonable predictor selection by removing irrelevant predictors and the predictors which just duplicate information from other more important predictors. The efficient selection of a predictor subset allows for finding the 'true' relationship between predictors and predictant, therefore avoiding the overfitting problem of ordinary multiple linear regression, especially when the size of potential predictors is large. MOS has been successfully applied and its algorithm has been described in detail in many studies (e.g. Paeth and Hense 2003; Paeth 2011). Here, we give a short summary.

MOS is trained with monthly outputs (including precipitation, temperature, and sea level pressure) from the CCLM assimilation run and E-OBS precipitation during 1961–2000. The training is not conducted for each month separately, but for a continuous record of monthly time series using complete simulation years. We conduct grid-point-wise bias-adjustment for monthly precipitation at the given model resolution over Europe. The predictand (y) is

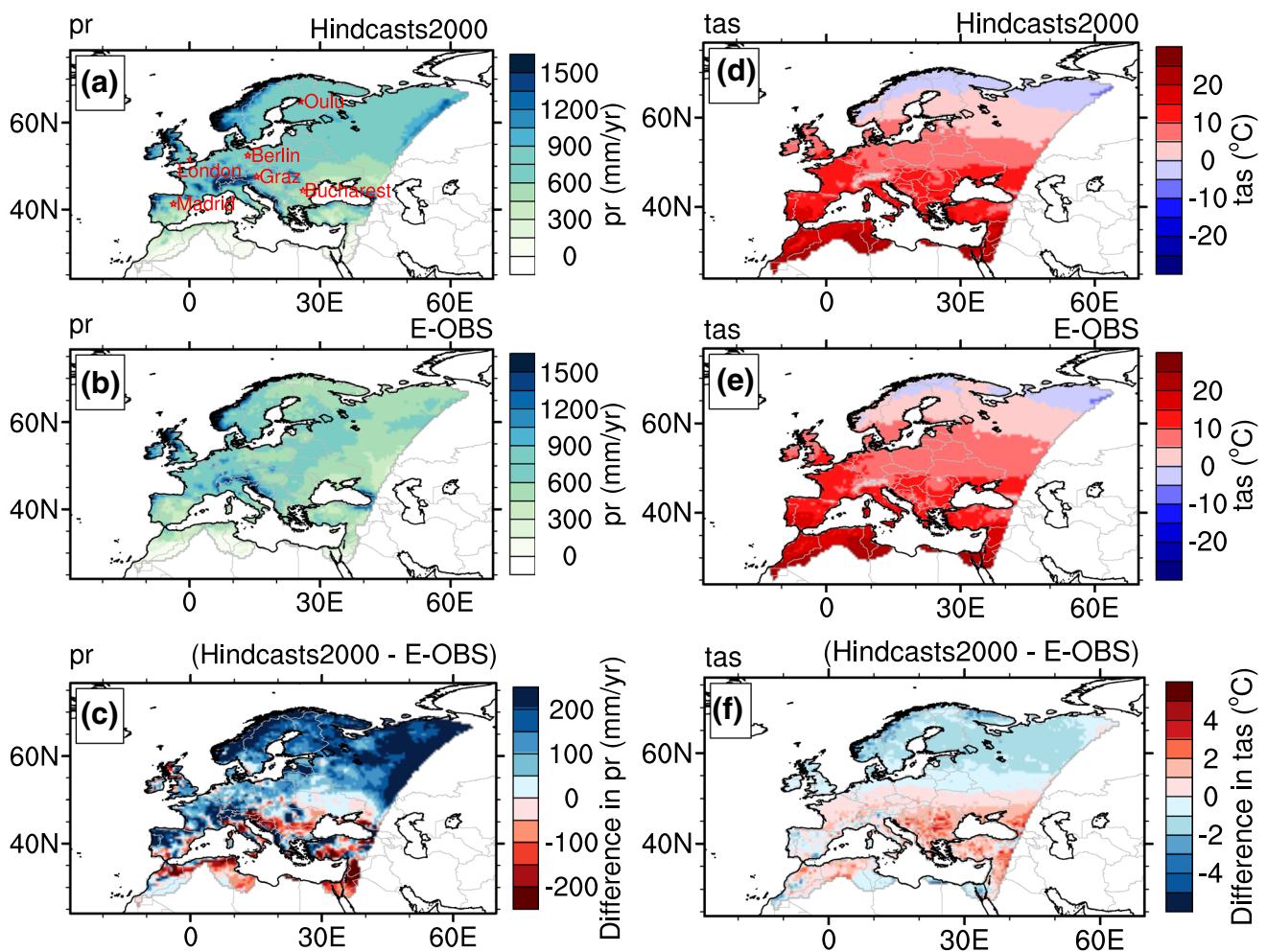


Fig. 1 Long-term mean annual precipitation (mm/year) (left column) and near-surface air temperature (°C) (right column) during 2001–2010 from CCLM decadal predictions (top row) and E-OBS

(middle row) and systematic differences between them (bottom row). Red stars and labels in sub plot **a** show the six example locations as discussed later in the paper

the observed monthly precipitation from the E-OBS data set. A total of 87 predictors (X) are prepared from simulated precipitation (pr), temperature (tas) and sea level pressure (psl). For each variable, two classes of predictors are defined: local and larger-scale empirical orthogonal function (EOF) predictors. For a given grid point, local predictors (pr_local, tas_local and psl_local) are selected from that grid point and its eight neighboring grid points. Thus, a potential spatial shift error of the model can be corrected. The complete model domain shares the same EOF predictors. The EOF predictors (pr_eof, tas_eof, psl_eof) are principal components (PC) determined from an empirical orthogonal function on the entire model domain. For each variable, 20 leading PC time series are taken into consideration (explaining $> 90\%$ of total variance of every predictor variable). Those EOF predictors represent larger-scale patterns of climate intra annual and innerannual climate variability in

Europe. In this study neither the predictant nor the predictors are standardized.

The predictors are selected stepwise according to the order of their bivariate correlation with predictand residuals. To avoid over-fitting, cross validation is used. The time series of predictors and predictand are split into two datasets: a training dataset (80%) on which the regression analysis is based on, and a control dataset (20%) to ensure the physical meaningfulness of the predictors. If an added predictor is reliable, the mean square error of both training and control dataset decreases. If not, the additional predictor enhances the forecast error (i.e. the mean square error of the control dataset). Therefore, the first minimum in the forecast error defines the optimal number of physically meaningful predictors in order of their importance. This procedure is iterated 100 times, each time selecting different training data by means of bootstrapping in a randomized way.

For a given time-period, the statistical relationship (so called *TF*) of k selected predictors ($x_{j=1,\dots,k}$) from the training simulation and the predictant y from the observational data is estimated via MOS using Eq. (1). Here, b_0 is the intercept term and $b_{j=1,\dots,k}$ are the regression coefficients of the selected predictors.

$$y = b_0 + \sum_{j=1}^k b_j x_j \quad (1)$$

The *TF* trained on the basis of the long-term CCLM assimilation run can then be transferred to each decadal hindcast member for bias-adjustment of the decadal predictions. The local predictors for the hindcasts are prepared in the same way as that for the 1961–2000 training simulation. The EOF predictors \overrightarrow{PC}_h of the hindcast predictors are projected on the EOFs from the training simulation. The idea behind this is that the predictor matrices of the training data and the decadal hindcasts are all projected onto the same spatial patterns. Using these selected predictors and the derived regression coefficients, the systematic error of CCLM can be adjusted for decadal prediction of precipitation with all other type of biases and prediction signal to be retained. As described in Sect. 2, the assimilation run limits the model errors to model's systematic error (e.g. over-/underestimation of rainfall). This model systematic error is embedded in the decadal predictions as well. Using hindcast directly, we derive another *TF* that adjusts the decadal precipitation bias in a different way based on the different type of predictors. Comparisons of the two TFs and their functionality and implications regarding the bias-adjustment for decadal prediction of precipitation are discussed in the following sections.

3.2 Prediction skill evaluations

We use the tool ‘Murphy–Epstein decomposition and Continuous Ranked Probability Skill Score’ (MurCSS) (Illing et al. 2014) to evaluate hindcasts’ precipitation prediction accuracy and uncertainty. This tool was developed for the evaluation of the MiKlip decadal prediction systems and has recently been applied in many studies (e.g. Kadow et al. 2015, 2017; Thoma et al. 2015; Timmreck et al. 2016; Pohlmann et al. 2017). In this study, we use MurCSS to compare the prediction performance of the bias-adjusted hindcasts against non-adjusted hindcasts as well as the climatological mean. Lead-time dependent evaluation (Boer et al. 2016) is considered. For the total hindcast period of 1961–2015, we use observations from years 1966–2015, 1967–2014 and 1961–2010 for evaluating lead-year 1–4, 2–5 and 6–9, respectively. MurCSS calculates two types of skill scores: (1) the Mean square error skill score (MSESS) and (2) the

Continuous ranked probability skill score (CRPSS). MSESS describes the accuracy of the hindcasts ensemble mean, positive/negative value of MSESS represent a gain/decrease of skill compared to the reference data. This score has two components: conditional bias and correlation. Correlation represents the linear relationship of hindcasts prediction and the observation. Conditional bias gives the value of correlation subtracting the ratio of standard deviation from hindcasts and observation. A positive value of conditional bias indicates a decrease of bias and vice versa. CRPSS on the other hand compares the uncertainty of the hindcasts ensemble spread with observations. CRPSS of 0 is the optimum, meaning that the uncertainty of prediction is adequate to the observation. Positive/negative values of CRPSS indicates an over-/underestimation of forecast uncertainty. The equation of these skill scores and their derivation can be found in detail in many publications (e.g., Goddard et al. 2013; Illing et al. 2014; Kadow et al. 2015).

This study focuses on testing whether bias-adjusted decadal hindcasts outperform the non-adjusted hindcasts in terms of prediction accuracy. In Sect. 5.1, we present MSESS ($H1/H2$, H , O) and its two components at a multi-year timescales, for the forecast years 1–4, 2–5 and 6–9. Here, $H1$ and $H2$ refers to our bias-adjusted decadal hindcasts, H represents the non-adjusted decadal hindcasts, and O is E-OBS. The skills in reference to the climatological mean (\bar{O}) of the non-adjusted hindcasts (H) and two bias-adjusted hindcasts ($H1/H2$), are provided in the supplementary material of this manuscript. Supplementary information includes MSESS ($H/H1/H2$, \bar{O} , O), correlation, conditional bias, and CRPSS for the forecast years 1–4, 2–5 and 6–9.

Besides the MurCSS evaluation at multi-yearly time-scale, we perform another approach of evaluation at the monthly time scale for a retrospective forecast period (Sect. 5.2). Since we use a historical time period of 1961–2000 for training of our statistical model, observations of later hindcasts are not touched and, therefore, the hindcast period 2001–2010 can be treated as a retrospective forecast. In this context, the month-to-month forecast variation is evaluated simply by comparing the estimated explained variances (R^2) and root mean square error (RMSE) between forecast and observation time series (120 monthly data-points for a 10-year decadal prediction period). R^2 and RMSE are estimated based on monthly time series of precipitation (not anomalies). Both non-adjusted and adjusted decadal prediction of precipitation are evaluated. This allows us to check how bias-adjustment changes the predicted monthly time series of decadal seasonal prediction. An increase of R^2 indicates an improved correlation between simulated and observed precipitation, this does not imply an enhanced climate predictability from month to month, and might suggest an improved representation of the seasonal cycle. Most

importantly, this analysis provides hints for linking the error sources to the functionality of our bias-adjustment. RMSE mainly measures the improvement of the simulated precipitation climatology.

4 Efficiency of the MOS approach in the CCLM assimilation run

In this section, we present results of the MOS training based on the assimilation run and some basic characteristics of the predictor selection. MOS is trained using predictors prepared from monthly outputs of CCLM assimilation run and E-OBS precipitation as predictand during 1961–2000. The monthly time series of each predictor and the predictand cover 480 data points (40 years \times 12 month). The derived multiple regression function, called *TF*, is estimated based on the raw values of those monthly time series. In the following, we discuss results from three aspects: (1) predictors selection, (2) the efficiency of MOS regression in terms of improved simulated precipitation and (3) the limitation of MOS regression.

4.1 MOS predictor selection

As a regularized regression method, MOS can identify a subset of predictors which are most related to the predictand among all considered predictors. As an example, predictors identified by MOS are illustrated in Fig. 2 for six model grid boxes across Europe. There is spatial variation in identified predictors for monthly precipitation. For example, only pr_local prevails for Berlin, while pr_local and psl_local are selected for Bucharest. A small proportion of the EOF predictors is selected occasionally for Graz, Madrid and London.

For an overview about the grid-point-wise identified predictors for Europe, we group similar predictors together. For example, all local predictors of precipitation are grouped as pr_local, all EOF predictors of precipitation are grouped as pr_eof (see x-labels in Fig. 2). We then calculated the percentage of number chosen within the six predictor types to the total number of selected predictors. For example, in total nine predictors are identified by MOS for London, among them eight are pr_local and one is pr_eof. Therefore, the percentage of selected pr_local is 88.9% ($8/9 \times 100$), the percentage of selected pr_eof is 11.1% ($1/9 \times 100$), and 0% for the other four predictor types. The spatial distribution of the

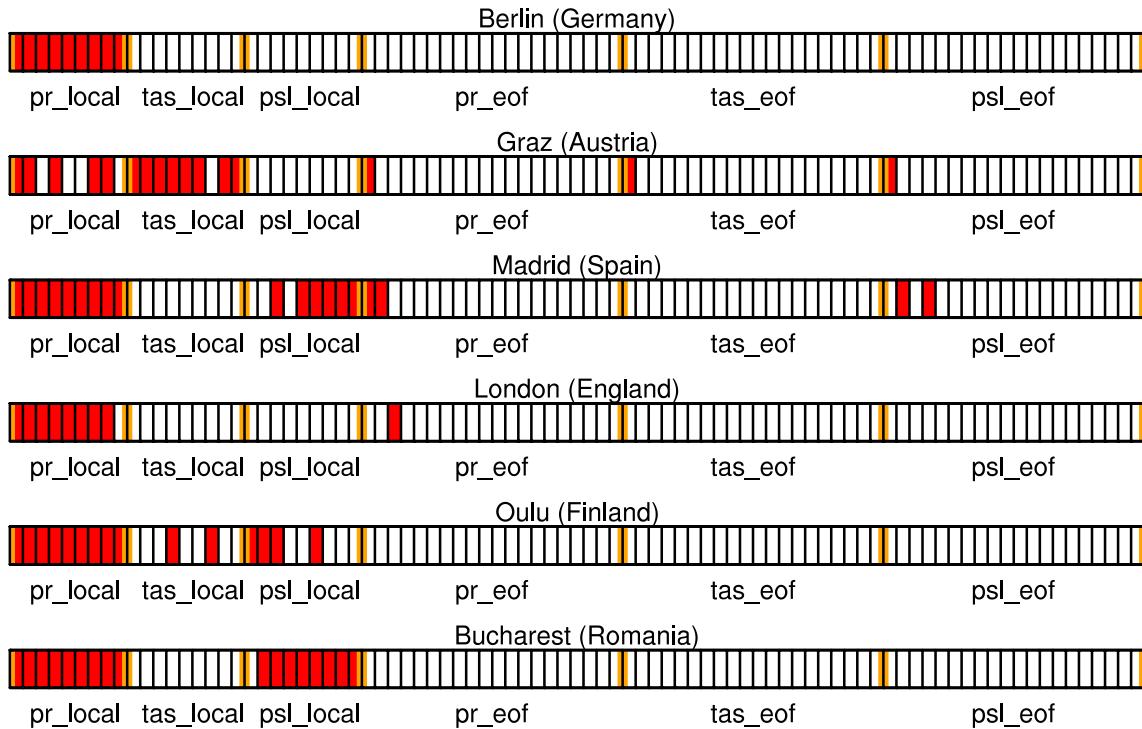


Fig. 2 MOS selection of predictors at six exemplary locations in Europe. MOS is trained with monthly data from outputs of the CCLM assimilation run (predictors) and E-OBS precipitation (predictand) during 1961–2000. 87 predictors in total (x-axis) are prepared from model outputs, including 9 local predictors and 20 EOF

predictors each for variables precipitation (pr), temperature (tas) and sea level pressure (psl). Red color indicates that the corresponding predictor is chosen by the MOS method. The groups of the six different predictor types (labels along the x-axis) are separated with orange bars

predictor selection for the six predictor types is displayed in Fig. 3. These results suggest that pr_local and psl_local predictors of the assimilation run are most useful for describing observed monthly precipitation variations. While pr_local is dominant for northern and western Europe (> 80% of all

selection), psl_local is important for middle eastern Europe (40–80%). tas_local plays a role in some regions in Poland, the Czech Republic and Northern Europe (10–40%). EOF predictors of the assimilation run appear to be less important and occupy a small proportion of less than 10%.

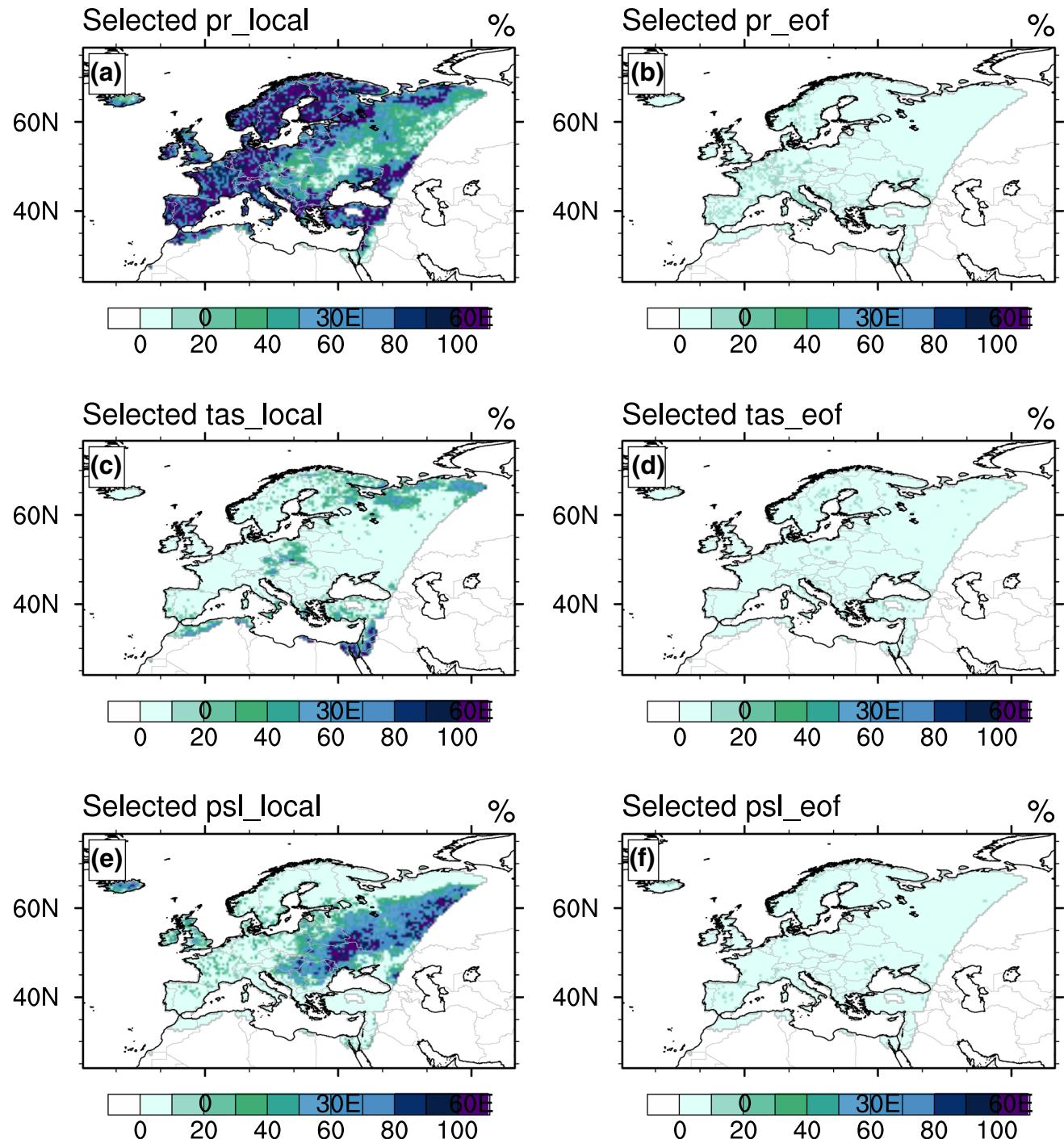


Fig. 3 Spatial distribution of MOS predictor selection for the six predictor groups. The MOS is trained with monthly data from CCLM assimilation outputs (predictors) and E-OBS precipitation (predictant)

during 1961–2000. The color bar shows the percentage of the respective predictor group with respect to the total number of selected predictors

4.2 Improvement in R^2 and reduction in RMSE by MOS regression

Figure 4 compares R^2 and RMSE for precipitation during the 40-year monthly time series between model and observations before MOS training, estimated precipitation after MOS training and their difference (after minus before). Being an assimilation run, the model has good skill in monthly precipitation estimation. It explains 40–80% of total variance for western and northern Europe, and 10–30% for middle and eastern Europe (Fig. 4a). This spatial pattern relates to the distribution of selected predictor types in Fig. 3. MOS chooses pr_local for regions where the simulated precipitation is most correlated with the observations, while for middle and eastern Europe, pr_local lost its priority and MOS select psl_local instead.

Note that this high explained variance between assimilation run and E-OBS cannot simply be explained by the seasonal cycle because the latter is not very marked over large parts of Europe—except for the Mediterranean basin where R^2 is not noticeably higher. MOS regression using selected predictors results in a substantial gain in the model skill for the whole domain (Fig. 4c, e). There is an increase in R^2 by 5–35%, higher for middle and eastern Europe than western and northern Europe (Fig. 4e). MOS regression decreases the model bias as well. The high values of RMSE (> 80 mm/month) for the eastern part of the model domain, the Alps and northwestern Scandinavian (Fig. 4b), are reduced to < 40 mm/month (Fig. 4d). Figure 4f shows an overall reduction in RMSE for Europe, mostly by 5–10 mm/month.

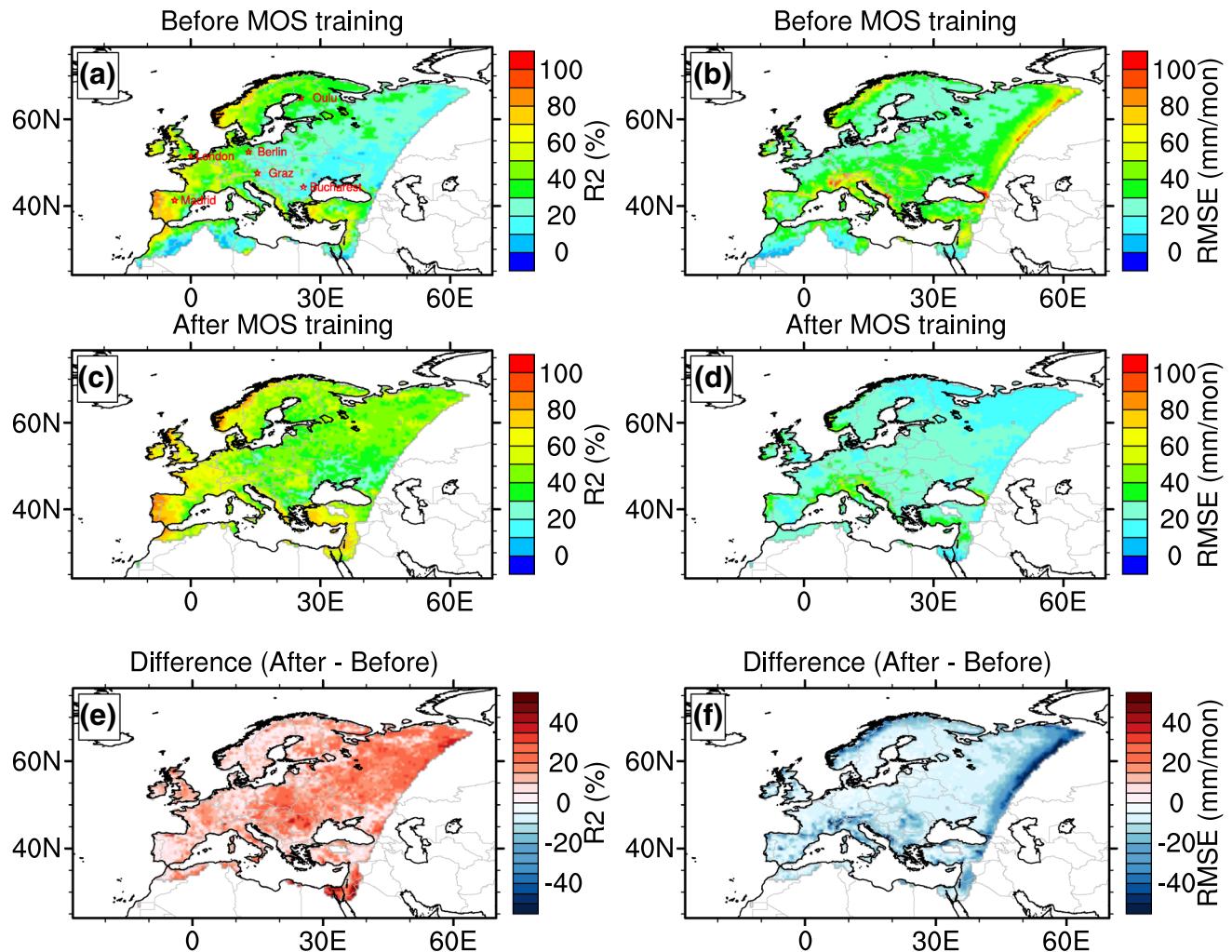


Fig. 4 Explained variance (R^2) (left column) and root mean square error (RMSE) (right column) of monthly precipitation during 1961–2000 between CCLM assimilation and E-OBS before the MOS-training (top raw), after MOS-training (middle raw) and the difference

(after–before) (bottom raw). Positive values in subplot (e) indicate that R^2 is improved by MOS regression. Negative values in subplot (f) indicate that RMSE is reduced by MOS regression

4.3 Reduced variances by MOS regression

Opposite to the gain in skill by MOS regression, it tends to underestimate month-to-month precipitation variability. This can be due to a reduced anomaly in monthly climatology or a reduced seasonal cycle. The standard derivation (SD) in 40-years monthly precipitation of three datasets (E-OBS, uncorrected and corrected CCLM assimilation run), are compared in Fig. 5. E-OBS monthly precipitation has a SD of 20–40 mm (Fig. 5a). The SD of the assimilated monthly precipitation agrees well with E-OBS both in its spatial distribution and amplitude, except for an overestimation along the eastern edge of the model domain (Fig. 5b). However, MOS regression reduces the SD of monthly precipitation for Europe. This results in an underestimation of precipitation variation by 5–10 mm (Fig. 5c). This is due to the fact that linear regression models primarily focus on the variables' centroid and underrepresent extreme anomalies. We will interpret the results of bias-adjustment for decadal hindcasts in the following section with this limitation in mind.

5 MOS approach applied to the decadal hindcasts/forecast

The last section has demonstrated MOS's efficiency in predictor selection and its skill in improving model accuracy. We found a *TF* which adjusts model output from the assimilation run towards observations. We now apply this *TF* to the 285 hindcasts runs (5 members \times 57 initializations) from CCLM individually for bias-adjustments. The bias-adjusted monthly precipitation is evaluated against observations at an aggregated multi-year time scale (Sect. 5.1) and at a monthly timescale for a retrospective forecast period (Sect. 5.2).

5.1 Multi-year time scale

After being initialized with a given observational state, decadal climate predictions tend to lose the observed climate variations and, thus, its predictive skill from one forecast year to the next. Therefore, decadal prediction is frequently evaluated at a lead-year dependent multi-year timescale (e.g. Kadow et al 2015, 2017). Here, we use the tool MurCSS developed specifically for evaluating the MiKlip decadal prediction system (Illing et al. 2014) (see details in Sect. 3.2), to evaluate our bias-adjustment on decadal hindcasts in a consistent way with other studies of the MiKlip project.

MSESS (together with conditional bias and correlation) of bias-adjusted hindcasts (H_1), with original hindcasts as reference (H), and measured against E-OBS (O), is analyzed for precipitation from all decadal hindcasts with start years varying between 1961 and 2014. The difference of MSESS from zero is tested at 5% of significance applying

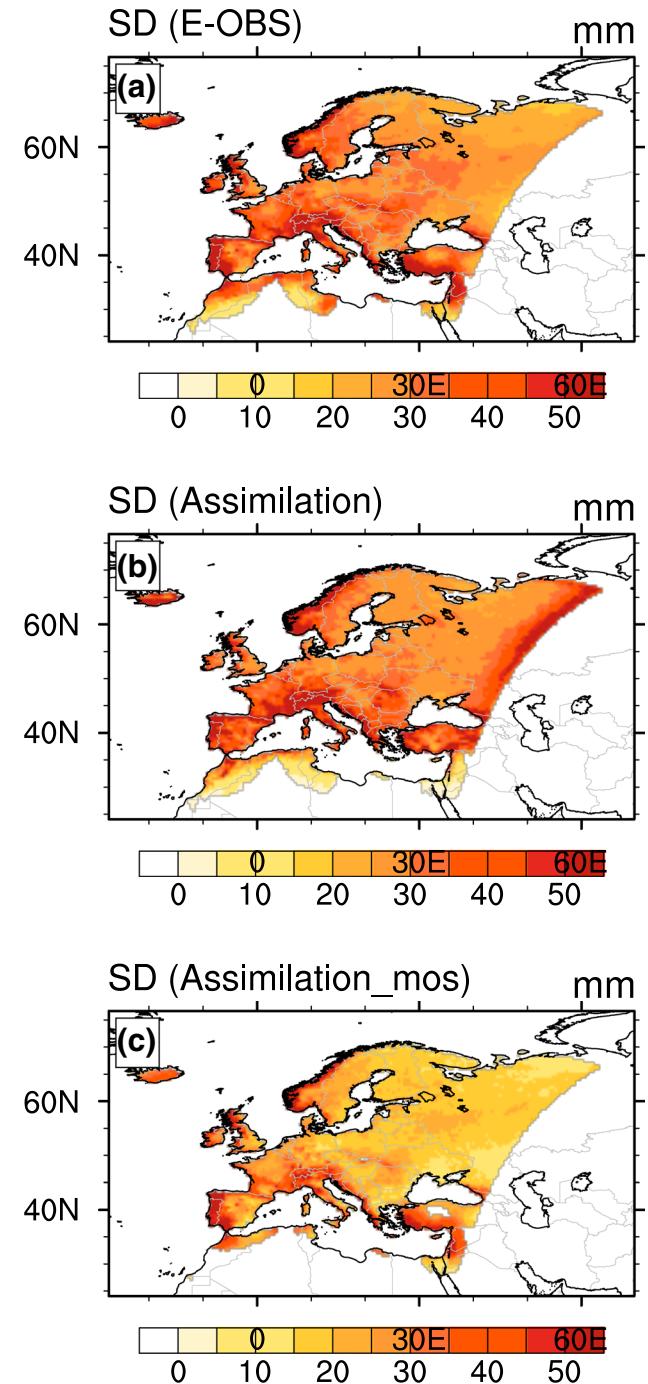


Fig. 5 Comparison between the standard deviation (SD) of 40-years monthly precipitation from E-OBS (a), from the uncorrected CCLM assimilation run (b) and the MOS-corrected CCLM assimilation run (c)

500 bootstraps. Figure 6 shows results for multi-year means of hindcast years 1–4, 2–5 and 6–9 using the *TF* derived from the 1961–2000 assimilation run (*TF1*). A systematic improvement in accuracy is observed regionally. Bias-adjustment with *TF1* seems to be more effective for some

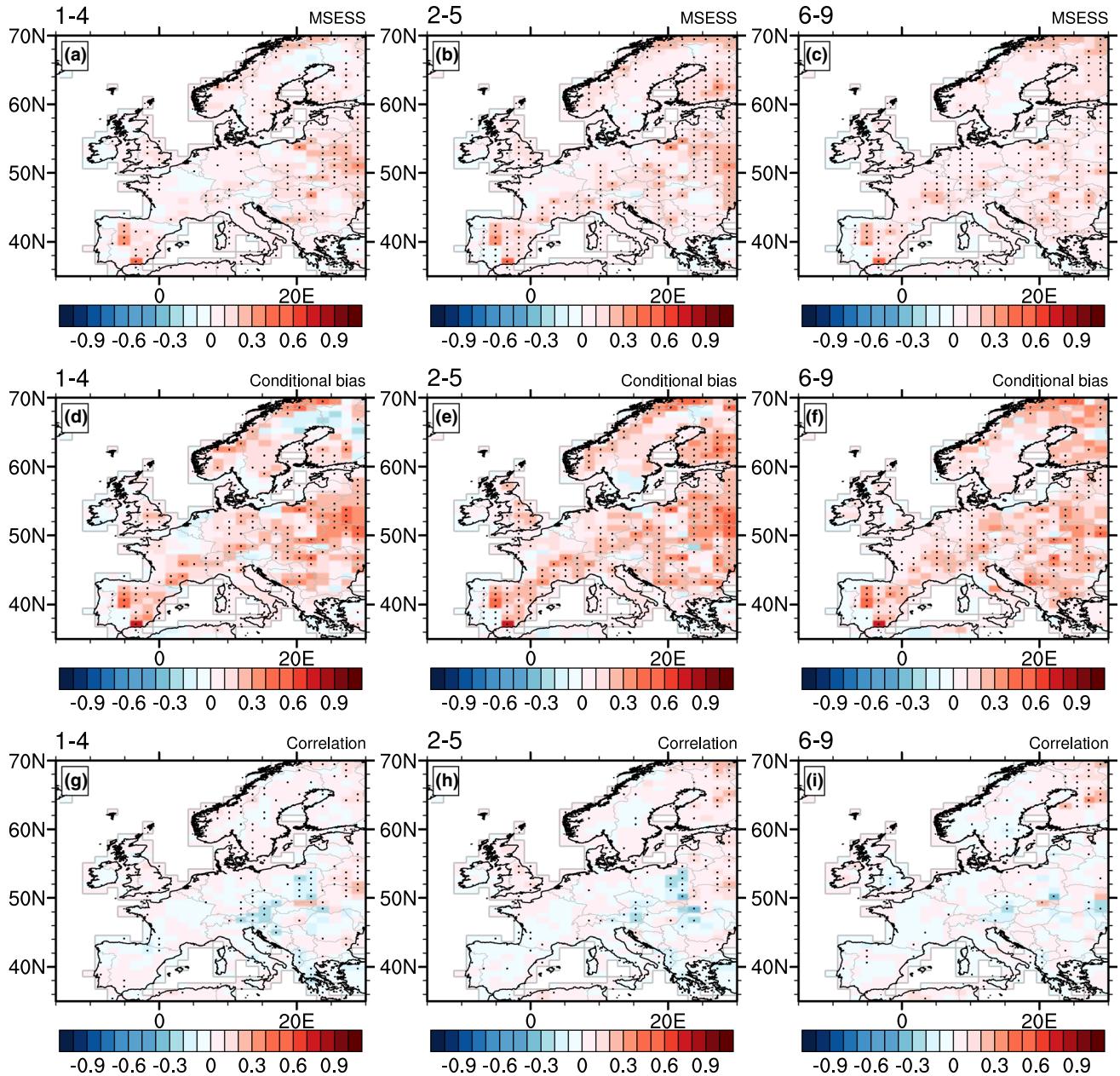


Fig. 6 Accuracy skill evaluation for forecast years 1–4 (left column), 2–5 (middle column) and 6–9 (right column) based on all MOS-corrected decadal hindcasts during the 1961–2014 period. Accuracy skill evaluation includes MSESS (H_1 , H , O) (top row), conditional bias (middle row) and correlation (bottom row). H_1 is the

TF_1 bias-adjusted hindcasts, H is the non-adjusted hindcasts, and O is the E-OBS during 1961–2014. The MOS TF refers to monthly time series trained on the CCLM assimilation run and E-OBS during 1961–2000. Black dots represent values significantly different from zero at 5% of significance applying 500 bootstraps

regions in the central Iberian Peninsula, northern and eastern Scandinavia, Ukraine and Belarus (MSESS up to 0.5) than for other regions in Europe (MSESS of 0.1–0.2). For the British Isles, TF_1 bias-adjustment results in an improved skill (mostly not significant) for hindcast year 6–9, and a slightly but significant increased skill (0.1–0.2) for hindcast year 1–4 and 2–5. Positive values of MSESS are accompanied by a decrease in the conditional bias (Fig. 6d–g) and

small changes (within ± 0.2) in correlation of hindcasts and observation (Fig. 6g–i). This indicates that the functionality of TF_1 adjustment is to reduce the hindcast's model's systematic error, and this adjustment has little influence on the trends or correlation of decadal precipitation prediction.

The above evaluations test if bias-adjusted hindcasts outperform the non-adjusted hindcasts in terms of precipitation prediction accuracy skill, but it might be interesting to assess

their skill scores in comparison with climatology or historical simulations as well. For this, we provide therefore supplementary material. Figure S1c–S3c in the supplementary material show that negative MSESS dominates in Europe before bias-adjustment. After *TF1* bias-adjustment, the size of regions with positive skill increase and prediction outperforms climatology in Europe (Fig. S1f–S3f). The skill gain in accuracy in reference to the climatology is accompanied with a regional underestimation in prediction uncertainty (Fig. S4). We see an adequate estimation for forecast uncertainty (CRPSS_{ES} of ~ 0) from the non-adjustment hindcast ensembles (Fig. S4a–c), but slightly underestimated uncertainty (CRPSS_{ES} of negative 0.05–0.1) for southeast Europe with our *TF1* bias-adjustment (Fig. S4d–f).

The strategy of using *TF1* from the CCLM assimilation run is to ensure a bias-adjustment only on the base of the model's systematic error. The assimilation run does not account for error sources due to initialization (initial shock, drift, and limited climate predictability at the seasonal to decadal time scales) Therefore, decadal hindcasts may be characterized by different systematic errors compared with the assimilation run. Indeed, previous decadal hindcast bias-adjustments are mostly based on the error evaluation of the hindcasts themselves (e.g. Kharin et al. 2012; Pasternack et al. 2018). To have a comparison, we run the MOS training with one arbitrary hindcast run (here for the 1986–1995 period) and the corresponding E-OBS data to produce a second *TF* (*TF2*) for bias-adjustment of all other decadal hindcasts. This also leads to a different selection of predictors (Fig. 7). Instead of the prevailing predictor types *pr_local* and *psl_local* as for the assimilation run (Fig. 3), *tas_local*, and EOF predictors are now selected. For eastern and northern Europe, *tas_local* is selected in more than 80% (Fig. 7c), *tas_eof* contributes the remaining part (Fig. 7d). This indicates that simulated temperature from the hindcasts is a useful predictor of the observed monthly precipitation changes in these regions. For western Europe, the largescale predictors (*pr_eof*, *tas_eof* and *psl_eof*) altogether represent more than 70% of all selected predictors (with $\sim 20\%$ each).

Using the new *TF2* derived from the hindcast run 1986–1995, we adjust and evaluate monthly precipitation from the complete set of decadal hindcasts again. The MSESS of *TF2* bias-adjusted hindcasts is shown in Fig. 8. This bias-adjustment achieves a larger improvement of decadal prediction skill than the bias-adjustment using *TF1* derived from the assimilation run (comparing Figs. 6 and 8). For the Iberian Peninsula, southern Finland, Ukraine and Belarus, the positive MSESS shown in Fig. 6 is approximately doubled (Fig. 8), accompanied by a decrease of conditional bias (up to 0.8) and increasing of correlation for this region (up to 0.4). On the other hand, this bias-adjustment results in an obvious reduction of skill in certain regions. For example, there is an MSESS of -0.2 to -0.4 in northern

Sweden and Finland for the hindcast years 1–4, and of -0.1 to 0.2 in some western Europe countries (e.g. France, Germany, Benelux countries) for hindcast years 2–5, where there is a decrease in correlation and an increased conditional bias.

In reference to the climatology, the *TF2* bias-adjustment gains more skill than the *TF1* bias-adjustment in terms of accuracy, positive skills dominate in Europe after *TF2* bias-adjustment (Fig. S1–S3). However, our *TF2* bias-adjustment causes a large amount of underestimation (mean CRPSS_{ES} of \sim negative 0.20–0.25) in prediction uncertainty (Fig. S4).

5.2 Monthly time scale

Decadal forecasting has been shown to have no seasonal skills after the first few forecast years (e.g. Stockdale et al. 2011; Gonzalez and Goddard 2016). However, improving forecast skill is the final goal of the developed decadal climate prediction system. Therefore, we still test our MOS approach in terms of the month-to-month precipitation variability within a retrospective forecast period (2001–2010). We do not expect a meaningful seasonal skill improvement for the whole 10 years' period. Rather, we use this analysis to discover how our bias-adjustment modifies the monthly time series and how are they related to model errors. More discussions are presented in Sect. 6. The transfer functions used in this subsection are derived once from the 1961–2000 CCLM assimilation run and once from the exemplary hindcast period 1986–1995. No observational information after 2000 has been taken into account during MOS training. Therefore, we can treat the decadal hindcast initialized in 2000 as a retrospective forecast and evaluate them against observations that are completely independent of bias adjustment.

We compare the R^2 and RMSE of monthly precipitation between the 2001–2010 hindcast and E-OBS for data before and after bias-adjustment. An example for decadal prediction bias-adjustment using *TF1* is illustrated in Fig. 9. Before bias-adjustment, the Hindcast 2001–2010 shows only small skill ($< 10\%$ in R^2) for most of the European region (Fig. 9a) and a pronounced bias ($> 100 \text{ mm/month}$ for western Scandinavia, the western Iberian Peninsula, the Alps and the eastern part of the model domain) (Fig. 9b). The low skill indicates that predicted precipitation is not closely related to observed monthly precipitation, which explains why MOS chooses *tas_local* and large-scale eof predictors instead of *pr_local* as predictors for observed precipitation when MOS is trained with a hindcast run (Fig. 7). After bias correction with *TF1*, the prediction shows a slight improvement in R^2 and a bias with smaller amplitudes (Fig. 9c, d). Here, bias-adjustment results in a less than 5% change in R^2 for most of the regions in Europe, except for an up to 25% improvement

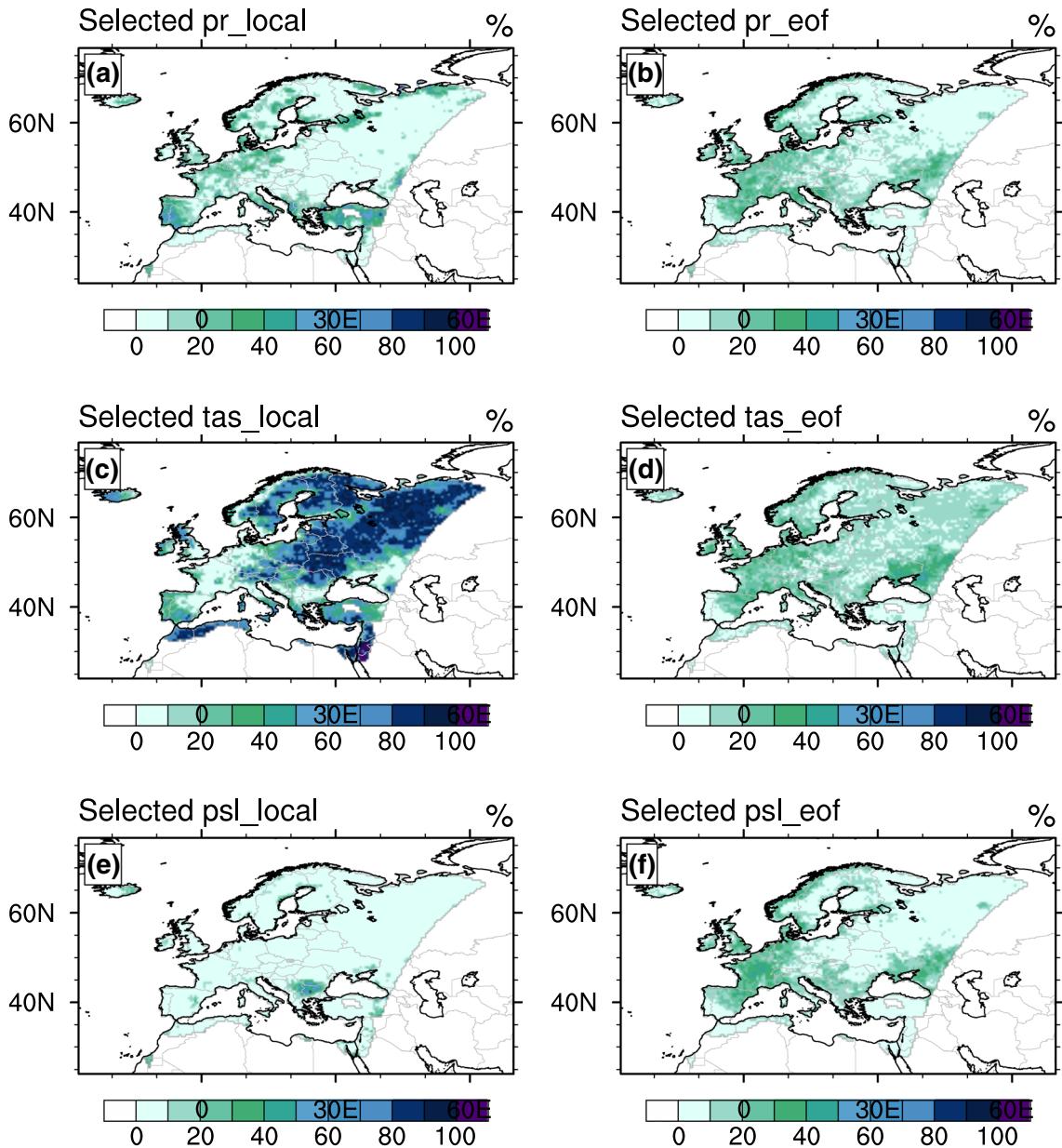


Fig. 7 Same as Fig. 3 but in this case the MOS trained with one exemplary decadal hindcast run from CCLM (period 1986–1995) using E-OBS as reference

for central Europe and up to 30% for northern Europe (Fig. 9e). Bias-adjustment with *TF1* reduces the monthly precipitation bias by 5–30 mm/month. The largest effect is found for northwestern Sweden, the Alps and the eastern part of the model domain (Fig. 9f).

For comparison, changes in R^2 and RMSE for the *TF2* bias-adjustment of the same hindcast 2001–2010 is shown in Fig. 10. The improvement in R^2 with *TF2* is ~1 to 15% higher than *TF1*, the RMSE reduction with *TF2* is about 1–10 mm/month higher than for *TF1* (compare Fig. 10a, b with Fig. 9e, f).

Figure 11 compares the 10-year monthly time series of precipitation at model grid boxes across Europe. We show bias-adjusted time series with *TF1* and *TF2*, the original hindcast and E-OBS during 2001–2010. Bias-adjustment with *TF1* mainly affects the amplitudes (compare blue lines with black lines) and hindcast month-to-month variability remains the same. That is caused by a bias-adjustment using local predictors. Instead, EOF predictors play a role when using *TF2*. For example, for Madrid (Fig. 11c), the bias-adjustment is calculated with diverse predictors including *pr_local*, *pr_eof*, *tas_eof* and *psl_eof* (Fig. 7). The adjusted

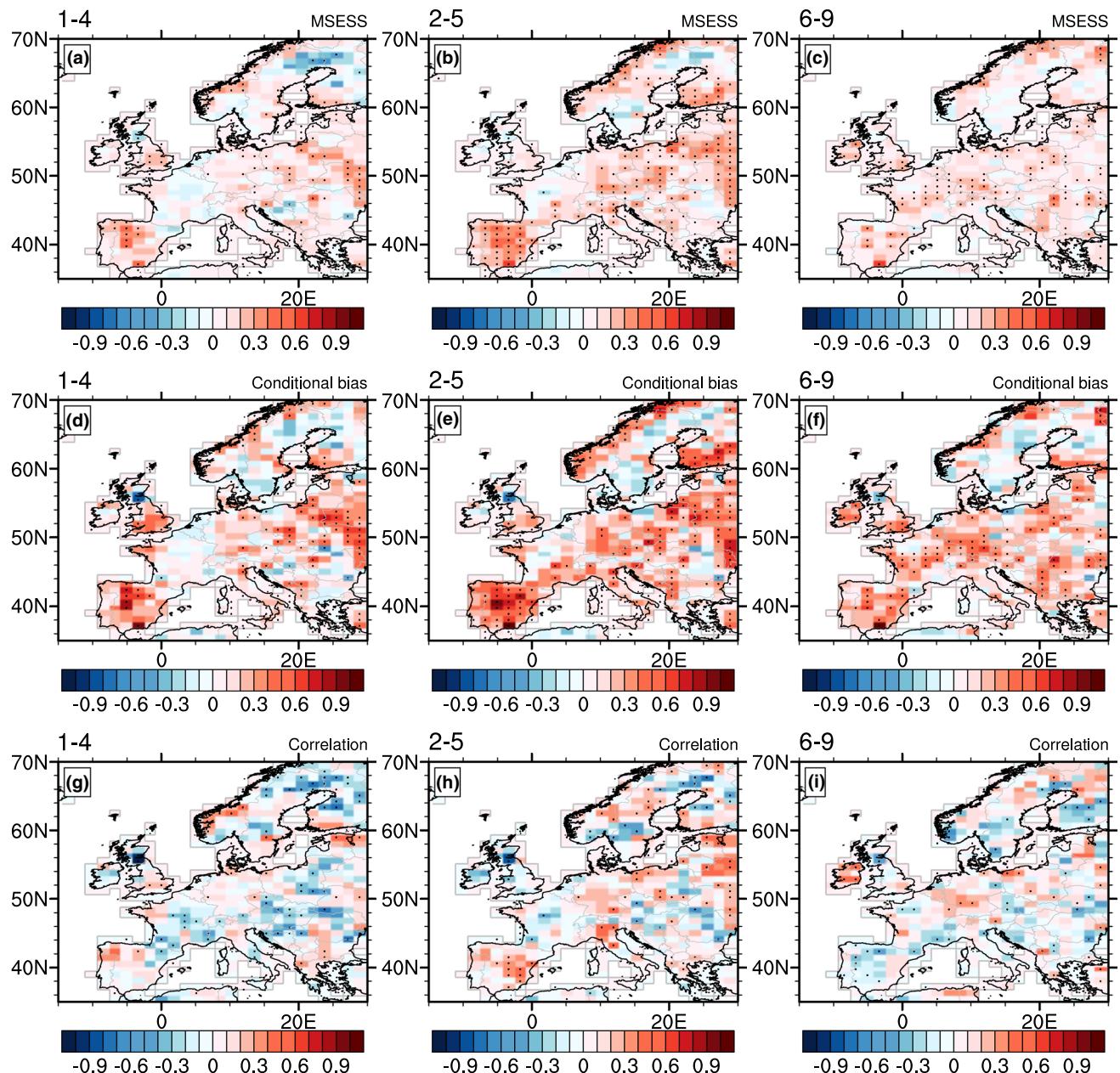


Fig. 8 Same as Fig. 6 but the MOS is trained with the 1986–1995 hindcast run

precipitation time-series (green line) are tuned more towards E-OBS (red line) than via the *TF1* adjustment (blue line).

6 Discussion

The MOS approach presented in this paper consists of two steps. In step one, a *TF* for precipitation is developed by training a long-term downscaled assimilation run with respect to observations. In step two, the derived *TF* is applied to a large set of decadal climate predictions. This

method has a wide application potential. It allows to identify and correct systematic model errors. This is specifically advantageous for initialized climate predictions because the risk of eliminating a part of the predictable climate signal is clearly reduced. The method can be applied to all climate variables and is not limited to a specific temporal nor spatial model resolution. The only criterion statistical model require is sufficient data for estimating a reliable *TF* between model output and observations. Based on our experience, we recommend to use time series with more than 100 data points in time.

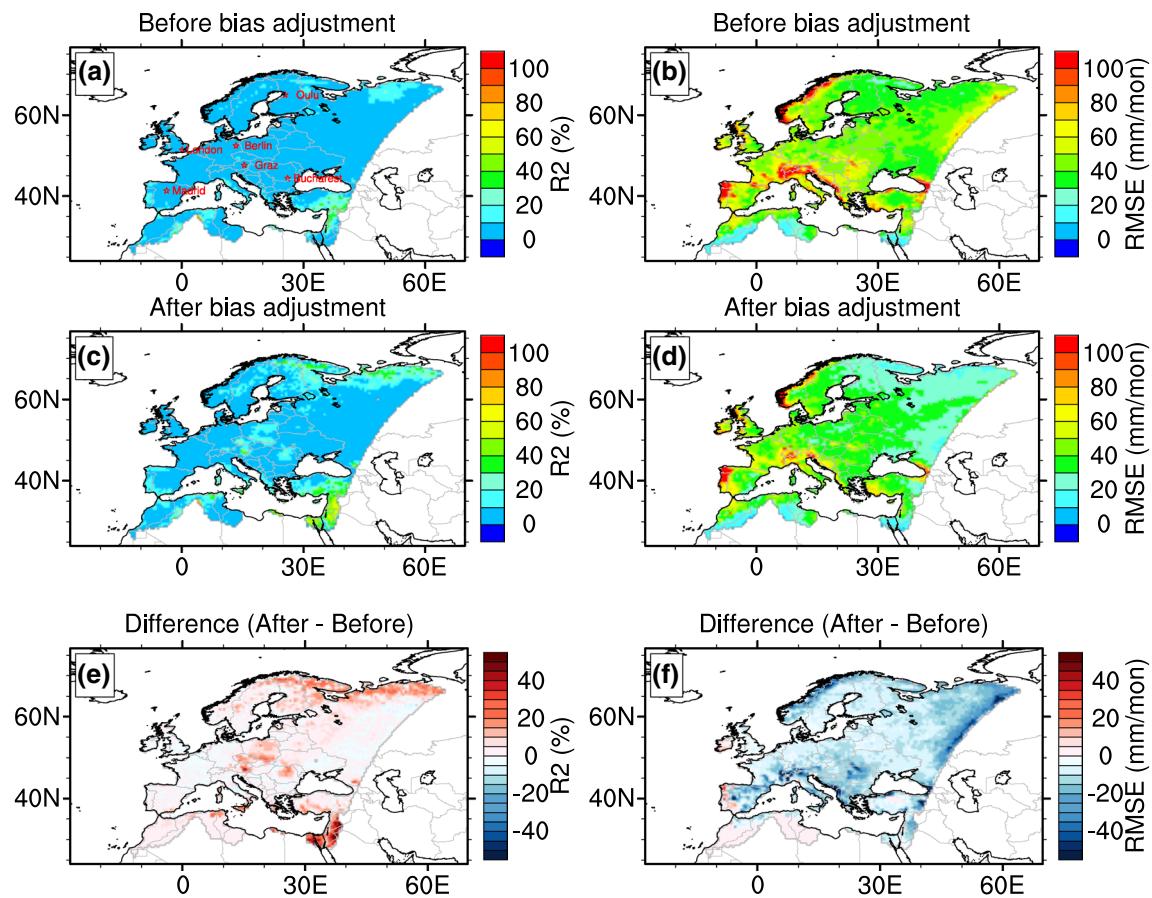


Fig. 9 Same as Fig. 4 but for one exemplary decadal hindcast period (2001–2010) with the MOS system being trained with the 1961–2000 CCLM assimilation run

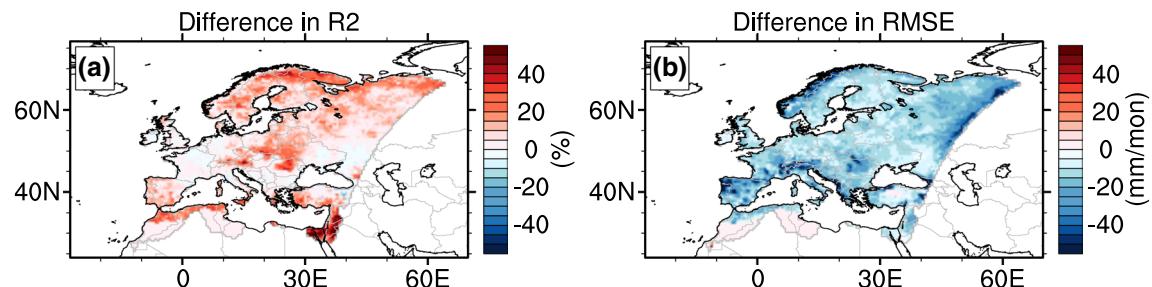
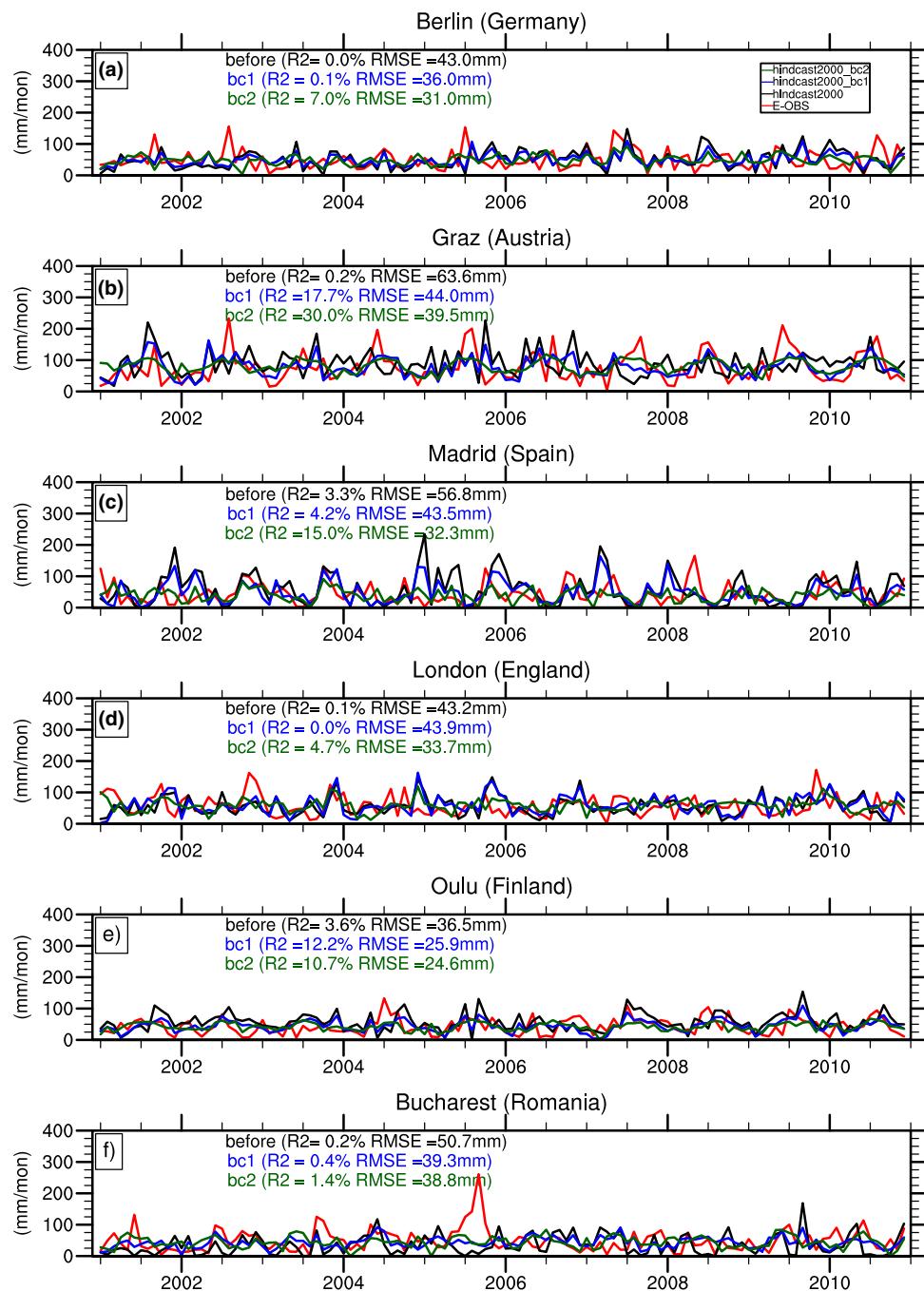


Fig. 10 Difference in R^2 and RMSE (after bias-adjustment–before bias-adjustment) due to MOS correction of one exemplary decadal hindcast period (2001–2010) with the MOS system being trained with the 1986–1995 decadal hindcast from CCLM

Our results demonstrate that the bias-adjustment on hindcasts/forecast is linked to the type of training simulation. Using an assimilation run (*TF1*) or a hindcast run (*TF2*) as training simulation makes a noticeable difference in bias-adjustment. *TF1* reduces the precipitation bias in the assimilation run, but does not substantially improve month-to-month variations in the decadal hindcasts. In comparison, bias reduction and correlation improvement with *TF2* are higher than with *TF1*, yet with some regional

differences: there are regions with significantly improved skill ($MSESS > 0.5$), as well as regions with reduced skill ($MSESS < -0.3$). These differences can be explained as follows: the assimilation run simulates the climate quite in phase with the observed climate system since it is assimilated with reanalysis data. The weather patterns and teleconnections of the assimilation experiment follow the observations. Therefore, the selected predictors, including local precipitation, are highly correlated with real precipitation

Fig. 11 Monthly precipitation time series (mm/month) during 2001–2010 at 6 exemplary locations across Europe from E-OBS (red lines), the uncorrected CCLM decadal hindcast (black line), the bias-adjusted CCLM decadal hindcast using *TF1* (blue lines) and using *TF2* (green lines). *TF1* refers to the 1961–2000 assimilation run and *TF2* to the 1986–1995 hindcast run. R^2 and RMSE between model and observations are indicated before and after MOS application



and the resulting *TF* yields precipitation time series with substantially increased R^2 and reduced RMSE. Applying the same *TF* to predictor time series from the decadal hindcasts, that suffer from additional errors arising from the model initialization and limited climate predictability, still removes a part of the systematic model bias but cannot push the bias corrected precipitation time series to the observed variations during the hindcast period (< 10% of R^2 in Fig. 9a). When training the MOS on the basis of a hindcast run, additional sources of discrepancy between model and observations can be removed with the risk of eliminating a part of the

predicted multi-year to decadal climate signal. Using EOF predictors allows to correct the state of major teleconnections, which are poorly predicted in the decadal predictions, but very important for the interannual climate variability. We have found also that the mean seasonal cycle of precipitation systematically differs between the assimilation run and the various decadal hindcasts on the other hand (not shown). Thus, training the MOS on the basis of a hindcast run also adjusts the mean seasonal cycle, leading to a higher correlation between observed and simulated monthly precipitation fluctuations (Fig. 10). However, note that this does not

implicitly enhance climate predictability from year to year or decade to decade (compare Fig. 10 with Fig. 8). Meanwhile, correcting all hindcasts monthly time series towards a known seasonal cycle removes variance from the datasets (e.g. Fig. 11) and results in a largely underestimated prediction uncertainty (Fig. S4h–j). This problem is due to that our MOS method can't distinguish the bias signal and the prediction signal within the hindcasts, and therefore an appropriate *TF2* which specifically addresses to the bias of the hindcasts cannot be estimated. Taken together, using *TF1* to adjust only the model systematic error for decadal prediction might be conservative and less effective, but allows the unusual but correct signals simulated for the future to be retained. The traditional way of training hindcasts with hindcasts can be recommended only if the applied statistical model can separate the prediction signal from bias and errors within the hindcast prediction. Clearly the MOS method is not capable to do so in its current setup.

Besides the choice of training simulations, other factors influence the bias-adjustment method as well. Previous studies suggest that bias-adjustment is sensitive to the chosen training period (e.g. Li et al. 2010) and statistical methods (e.g. Teutschbein and Seibert 2012; Piani and Haerter 2012; Cannon 2016). To address the first point, we conducted a series of sensitivity tests with various training time periods for both types of our training simulation (plots not shown). We tested five different 30-years training periods with a shift of 5 years from the assimilation run and used another time-period (1976–1985) as reference hindcast run. The shifts of training period for the assimilation run results in a minor difference in the derived *TFs*, indicating that 30-years of climatological information are sufficient to estimate model errors. In contrast, decadal hindcasts cover only a 10-year period, hence the *TFs* are more sensitive to the choice of training period. But the magnitude of differences due to sample size and training time-period is much less than that caused by the difference in the type of the training simulation. As for the influence of statistical methods, we will have a thorough discussion in a succeeding paper.

The flexible grid box-wise selection of predictors is an especially useful feature of this method, as climate models have different skills from region to region and from variable to variable. For example, MOS trained with the assimilation run chooses *psl_local* instead of *pr_local* predictors for central and eastern Europe. This is because CCLM's skill in precipitation prediction is weak in this region and observations can be better estimated using the simulated sea level pressure. Another example is that MOS trained with a hindcast run selects *tas_local* and EOF predictors for most parts of Europe. Hence, bias-correction of precipitation is not dependent on simulated precipitation itself, but on other thermodynamic variables. This multi-variate approach retains the dynamic link between the model's variables, in

contrast to a 1D bias-adjustment (e.g. precipitation bias-adjustment with precipitation). The advantage of multi-dimensional bias-adjustment has already been demonstrated by previous studies (e.g. Piani and Haerter 2012; Cannon 2016).

The underestimated precipitation variability is a limitation of this method, especially when it comes to the prediction of extreme events. Our bias-correction tends to suppress extreme values (Figs. 5, 11) which results in overestimated amount of weak and mid-range anomalies and in an underestimation of the forecast uncertainty (Fig. S4 in supplementary material). However, a proper estimation of uncertainty is important for forecasts. Therefore, when applying this method, we should bear in mind that the real climate variability is larger than the bias-adjusted value. Further investigation and algorithm development are needed to overcome the above shortcoming.

7 Conclusions

In this study, we statistically calibrate simulated precipitation from decadal climate prediction runs with the high-resolution regional climate model CCLM over Europe. We developed a MOS system that determines *TFs* based on two different types of CCLM reference simulations and E-OBS data. There are three main conclusions to be drawn from this study: (1) the presented MOS approach is capable of selecting predictors relevant to observations, of improving simulation skill, and of reducing the systematic model bias. However, MOS has a limitation in that it underestimates the observed precipitation variability. (2) Results of bias-adjustment in decadal hindcasts/forecasts are promising. MSESS evaluation at the multi-year timescale demonstrates that our bias adjustment achieves an increase of regionally more than 0.5. (3) The *TF2* bias-adjustment is more effective than the *TF1* one in terms of the decadal prediction accuracy skill, but that is accompanied with an underestimation of forecast uncertainty and an elimination of prediction signals. This problem is caused by the fact that our current statistical method cannot distinguish information from the hindcast training dataset such as useful prediction and bias/error. *TF1* on the other hand, is estimated specifically for model systematic error adjustment. Although *TF1* bias-adjustment for hindcasts prediction is not exclusive, it leads to a least underestimation in climate prediction uncertainty.

We suggest that for variables like precipitation which shows generally poor skills in its decadal prediction and therefore the lead-time-depend bias might not be dominant, it is more conservative and stringent to derive systematic model biases from assimilation runs, but this allows to retain potentially predictable climate signal in decadal hindcast and forecasts. Using hindcasts itself for training is recommend

only when a statistical method can distinguish specifically biases and climate predictions within a hindcast dataset.

Acknowledgements This work was conducted in the framework of the German MiKlip project and supported by the German Ministry of Education and Research (BMBF) under Grant no. 01LP1129A-F. We thank the Max-Planck Institute for Meteorology for providing the MPI-ESM decadal predictions. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>). Support for the Twentieth Century reanalysis Project dataset is provided by the US Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) Program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. We thank the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing Ocean Reanalysis, ERA-20C and ERA-Interim datasets. All MiKlip simulations were performed at the German climate computing center (Deutsches Klimarechenzentrum, DKRZ).

References

- Balmaseda MA et al (2015) The ocean reanalyses intercomparison project (ORA-IP). *J Oper Oceanogr*. <https://doi.org/10.1080/175576X.2015.1022329>
- Boer GJ, Smith DM, Cassou C, Doblas-Reyes F, Danabasoglu G, Kirtman B, Kushner Y, Kimoto M, Meehl GA, Msadek R, Mueller WA, Taylor KE, Zwiers F, Rixen M, Ruprich-Robert Y, Eade R (2016) The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci Model Dev* 9:3751–3777. <https://doi.org/10.5194/gmd-9-3751-2016>
- Buizza R, Richardson D (2017) 25 years of ensemble forecasting at ECMWF. *ECMWF Newsl* 153:20–31. <https://doi.org/10.21957/bv418o>
- Cannon A (2016) Multivariate bias correction of climate model output: matching marginal distributions and intervariable dependence structure. *J Clim* 29:7045–7064. <https://doi.org/10.1175/JCLI-D-15-0679.1>
- Dee DP et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137:553–597. <https://doi.org/10.1002/qj.828>
- Fuckar NS, Volpi D, Guemas V, Doblas-Reyes FJ (2014) A posteriori adjustment of near-term climate predictions: accounting for the drift dependence on the initial conditions. *Geophys Res Lett* 41:5200–5207. <https://doi.org/10.1002/2014GL060815>
- Giorgetta MA, Jungclaus J, Reick CH, Legutke S, Bader J, Boettinger M, Brovkin V, Crueger T, Esch M, Fieg K, Glushak K, Gayler V, Haak H, Hollweg H-D, Ilyina T, Kinne S, Kornblueh L, Matei D, Mauritsen T, Mikolajewicz U, Mueller W, Notz D, Pithan F, Raddatz T, Rast S, Redler R, Roeckner E, Schmidt H, Schnur R, Segschneider J, Six KD, Stockhause M, Timmreck C, Wegner J, Widmann H, Wieners K-H, Claussen M, Marotzke J, Stevens B (2013) Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5: climate changes in MPI-ESM. *J Adv Model Earth Syst* 5:572–579. <https://doi.org/10.1002/jame.20038>
- Goddard L et al (2013) A verification framework for interannual-to-decadal predictions experiments. *Clim Dyn* 40:245–272. <https://doi.org/10.1007/s00382-012-1481-2>
- Gonzalez PLM, Goddard L (2016) Long-lead ENSO predictability from CMIP5 decadal hindcasts. *Clim Dyn* 46:3127–3147. <https://doi.org/10.1007/s00382-015-2757-0>
- Hawkins E, Smith RS, Gregory JM, Stainforth DA (2016) Irreducible uncertainty in near-term climate projections. *Clim Dyn* 46:38–47. <https://doi.org/10.1007/s00382-015-2806-8>
- Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M (2008) A European daily high-resolution gridded dataset of surface temperature and precipitation. *J Geophys Res* 113:D20119. <https://doi.org/10.1029/2008JD10201>
- Hudson D, Alves O, Hendon HH, Marshall AG (2011) Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia. *Q J R Meteorol Soc* 137:673–689
- Illing S, Kadow C, Kunst O, Cubasch U (2014) MurCSS: a tool for standardized evaluation of decadal hindcast systems. *J Open Res Softw* 2:e24. <https://doi.org/10.5334/jors.bf>
- Jungclaus JH, Fischer N, Haak H, Lohmann K, Marotzke J, Matei D, Mikolajewicz U, Notz D, von Storch JS (2013) Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model: MPIOM CMIP5 Ocean Simulations. *J Adv Model Earth Syst* 5:422–446. <https://doi.org/10.1002/jame.20023>
- Kadow C, Illing S, Kunst O, Rust HW, Pohlmann H, Mueller WA, Cubasch U (2015) Evaluation of forecasts by accuracy and spread in the MiKlip decadal climate prediction system. *Meteorol Z* 25:631–643. <https://doi.org/10.1127/metz/2015/0639>
- Kadow C, Illing S, Kroener I, Ulbrich U, Cubasch U (2017) Decadal climate predictions improved by ocean ensemble dispersion filtering. *J Adv Model Earth Syst* 9:1138–1149. <https://doi.org/10.1002/2016MS000787>
- Kharin VV, Merryfield GJ, Merryfield WJ, Scinocca JF, Lee W-S (2012) Statistical adjustment of decadal predictions in a changing climate. *Geophys Res Lett* 39:L19705. <https://doi.org/10.1029/2012GL052647>
- Kirshnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Williford CE, Gadgil S, Surendran S (1999) Improved weather and seasonal climate forecasts from multimodel superensemble. *Nature* 285:1548–1550
- Kruschke T, Rust HW, Kadow C, Mueller WA, Pohlmann H, Leckebusch GC, Ulbrich U (2015) Probabilistic evaluation of decadal prediction skill regarding Northern Hemisphere winter storms. *Meteorol Z* 25:721–738
- Li H, Sheffield J, Wood EF (2010) Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J Geophys Res* 115:D10101. <https://doi.org/10.1029/2009JD012882>
- Marotzke J, Mueller WA, Vamborg FSE, Becker P, Cubasch U, Feldmann H, Kaspar F, Kottmeier C, Marini C, Polkova I, Proemmel K, Rust HW, Stammer D, Ulbrich U, Kadow C, Koehl A, Kroeger J, Kruschke T, Pinto JG, Pohlmann H, Reyers M, Schroeder M, Sienz F, Timmreck C, Ziese M (2016) MiKlip—a national research project on decadal climate prediction. *Bull Am Meteorol Soc*. <https://doi.org/10.1175/BAMS-D-15-00184.1>
- Meehl GAL et al (2014) Decadal climate prediction: an update from the trenches. *Bull Am Meteorol Soc* 95:243–267. <https://doi.org/10.1175/BAMS-D-12-00241.1>
- Mieruch S, Feldmann H, Schaedler S, Lenz C-J, Kothe S, Kottmeier C (2014) The regional MiKlip decadal forecast ensemble for Europe: the added value of downscaling. *Geosci Model Dev* 7:2983–29999. <https://doi.org/10.5194/gmd-7-2983-2014>
- Müller WA, Baehr J, Haak H, Jungclaus JH, Kroeger J, Matei D, Notz D, Pohlmann H, von Storch J-S, Marotzke J (2012) Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Plank Institute for Meteorology. *Geophys Res Lett* 39:L22707. <https://doi.org/10.1029/2012GL053326>
- Müller WA, Jungclaus JH, Mauritsen T, Baehr J, Bittner M, Budich R, Bunzel F, Esch M, Ghosh R, Haak H, Ilyina T, Kleine T, Kornblueh L, Li H, Modali K, Notz D, Pohlmann H, Roeckner E,

- Stemmler I, Tian F, Marotzke J (2018) A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *J Adv Model Earth Syst.* <https://doi.org/10.1029/2017MS001217>
- Paeth H (2011) Postprocessing of simulated precipitation for impact research in West Africa. Part I: model output statistics for monthly data. *Clim Dyn* 36:1321–1336. <https://doi.org/10.1007/s00382-010-0760-z>
- Paeth H, Hense A (2003) Seasonal forecast of sub-sahelian rainfall using cross validated model output statistics. *Meteorol Z* 12:157–193. <https://doi.org/10.1127/0941-2948/2003/0012-0157>
- Paeth H, Paxian A, Sein DV, Jacob D, Panitz H-J, Warscher M, Fink AH, Kunstmann H, Breil M, Engel T, Krause A, toedter J, Ahrens B (2017) Decadal and multi-year predictability of the West African monsoon and the role of dynamical downscaling. *Meteorol Z.* <https://doi.org/10.1127/metZ/2017/0811>
- Paeth H, Li J, Pollinger F, Mueller WA, Pohlmann H, Feldmann H, Panitz H-J (2018) An effective drift correction for dynamical downscaling of decadal global climate predictions. *Clim Dyn.* <https://doi.org/10.1007/s00382-018-4195-2>
- Panitz H-J, Fosser G, Sasse R, Sedlmeier K, Mieruch S, Breil M, Feldmann H, Schaedler G (2014) High resolution climate modelling with the CCLM regional model. *High Perform Comput Simul USC Comput Sci.* <https://doi.org/10.1007/978-3-319-02165-2>
- Pasternack A, Bhend J, Liniger MA, Rust HW, Mueller WA, Ulbrich U (2018) Parametric decadal climate forecast recalibration (DeFoReSt 1.0). *Geosci Model Dev* 11:351–368. <https://doi.org/10.5194/gmd-11-351-2018>
- Pattatanus-Abraham M, Kadow C, Illing S, Mueller W, Pohlmann H, Steinbrecht W (2016) Bias and drift pf the mid-range decadal climate prediction system (MiKlip) validated by European radiosonde data. *Meteorol Z* 25:709–720. <https://doi.org/10.1127/metZ/2016/0803>
- Paxian A et al (2016) Bias reduction in decadal predictions of West African monsoon rainfall using regional climate models. *J Geophys Res.* <https://doi.org/10.1002/2015JD024143>
- Piani C, Haerter JO (2012) Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophys Res Lett* 39:L20401. <https://doi.org/10.1029/2012GL053839>
- Pohlmann H, Mueller WA, Kulkarni K, Kameswarao M, Matei D, Vamborg FSE, Kadow C, Illing S, Marotzke J (2013) Improved forecast skill in the tropics in the new MiKlip decadal climate predictions. *Geophys Res Lett* 40:5798–5802
- Pohlmann H, Kroeger J, Greatbatch RJ, Mueller WA (2017) Initialization shock in decadal hindcasts due to errors in wind stress over the tropical Pacific. *Clim Dyn* 49:2685–2693. <https://doi.org/10.1007/s00382-016-3486-8>
- Reyers M, Pinto JG, Moemken J (2015) Statistical-dynamical downscaling for wind energy potentials: evaluation and applications to decadal hindcasts and climate change projection. *Int J Climatol* 35:229–244. <https://doi.org/10.1002/joc.3975>
- Rockel B, Will A, Hense A (2008) The regional climate model COSMO-CLM (CCLM). *Meteorol Z* 17:347–348
- Sansom PG, Ferro CAT, Stephenson DB, Goddard L, Mason SJ (2016) Best practices for postprocessing ensemble climate forecast. Part I: selecting appropriate recalibration methods. *J Clim* 29:7247–7264
- Smith LA, Du H, Suckling EB, Niehoerster F (2015) Probabilistic skill in ensemble seasonal forecasts. *Q J R Meteorol Soc* 141:1085–1100. <https://doi.org/10.1002/qj.2403>
- Stockdale TN, Anderson DLT, Balmaseda MA, Doblas-Reyes F, Ferranti L, Mogensen K, Palmer TN, Molteni F, Vitart F (2011) ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Clim Dyn* 37:455–471. <https://doi.org/10.1007/s00382-010-0947-3>
- Teutschbein C, Seibert J (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. *J Hydrol* 456–457:12–29
- Thoma M, Greatbatch RJ, Kadow C, Gerdes R (2015) Decadal hindcasts initialized using observed surface wind stress: evaluation and prediction out to 2024. *Geophys Res Lett* 42:6454–6461. <https://doi.org/10.1002/2015GL064833>
- Timmreck C, Pohlmann H, Illing S, Kadow C (2016) The impact of stratospheric volcanic aerosol on decadal-scale climate predictions. *Geophys Res Lett* 43:834–842. <https://doi.org/10.1002/2015GL067431>
- Uppala SM, Källberg PW, Simmons AJ, Andrae U, Bechtold VdaC, Florino M, Gribson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo R, Woollen J (2005) The ERA-40 re-analysis. *Q J R Meteorol Soc* 132:2961–3012. <https://doi.org/10.1256/qj.04.176>
- WMO (2011) Climate knowledge for action: a global framework for climate services-Empowering the most vulnerable. WMO Rep 1065:247

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.