



The regional MiKlip decadal forecast ensemble for Europe: the added value of downscaling

S. Mieruch¹, H. Feldmann¹, G. Schädler¹, C.-J. Lenz², S. Kothe³, and C. Kottmeier¹

¹Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany

²Deutscher Wetterdienst, Offenbach, Germany

³Institute for Atmospheric and Environmental Sciences, Goethe-University Frankfurt, Frankfurt am Main, Germany

Correspondence to: S. Mieruch (sebastian.mieruch@kit.edu)

Received: 30 September 2013 – Published in Geosci. Model Dev. Discuss.: 22 November 2013

Revised: 18 October 2014 – Accepted: 31 October 2014 – Published: 17 December 2014

Abstract. The prediction of climate on time scales of years to decades is attracting the interest of both climate researchers and stakeholders. The German Ministry for Education and Research (BMBF) has launched a major research programme on decadal climate prediction called MiKlip (Mittelfristige Klimaprognosen, Decadal Climate Prediction) in order to investigate the prediction potential of global and regional climate models (RCMs). In this paper we describe a regional predictive hindcast ensemble, its validation, and the added value of regional downscaling. Global predictions are obtained from an ensemble of simulations by the MPI-ESM-LR model (baseline 0 runs), which were downscaled for Europe using the COSMO-CLM regional model. Decadal hindcasts were produced for the 5 decades starting in 1961 until 2001. Observations were taken from the E-OBS data set. To identify decadal variability and predictability, we removed the long-term mean, as well as the long-term linear trend from the data. We split the resulting anomaly time series into two parts, the first including lead times of 1–5 years, reflecting the skill which originates mainly from the initialisation, and the second including lead times from 6–10 years, which are more related to the representation of low frequency climate variability and the effects of external forcing. We investigated temperature averages and precipitation sums for the summer and winter half-year. Skill assessment was based on correlation coefficient and reliability. We found that regional downscaling preserves, but mostly does not improve the skill and the reliability of the global predictions for summer half-year temperature anomalies. In contrast, regionalisation improves global decadal predictions of half-year precipitation sums in

most parts of Europe. The added value results from an increased predictive skill on grid-point basis together with an improvement of the ensemble spread, i.e. the reliability.

1 Introduction

Interest in longer-term climate predictions in the range of about 10 years is growing. Such predictions, as opposed to projections that do not take into account the influence of initial conditions, would be very useful for all branches of public life and for planning purposes, e.g. in agriculture, energy management, hydrology, and health. In the sense of seamless predictions (Palmer et al., 2008), a decadal prediction system would well complement existing short range systems, as well as seasonal predictions provided by ECMWF (<http://www.ecmwf.int/products/changes/system4/>) and CPC (<http://www.cpc.ncep.noaa.gov/products/predictions/90day/>), for instance. Decadal climate predictions present a major scientific challenge. It is not known yet to what extent useful predictions are possible in terms of lead time, geographical position, spatial resolution, meteorological variables, and statistics such as means or extremes. There is, however, widespread agreement about the necessary requirements for such predictions to be successful: (i) coupled ocean-atmosphere models are likely the most effective means for making global climate predictions (and projections); (ii) predictability lies mainly in the slow components of the climate system, i.e. oceans, sea ice, soil (Chikamoto et al., 2014), and atmospheric processes such as

the quasi-biennial oscillation (QBO) (Scaife et al., 2014b) and the North Atlantic oscillation (NAO) (Scaife et al., 2014a). Skilful modelling and good initialisation of these components is essential (Keenlyside et al., 2008); (iii) predictability must come from the large-scale processes and interactions, such as those of the Atlantic multidecadal oscillation (AMO), El Niño–Southern Oscillation (ENSO), QBO, and NAO, which must be captured by the global models; (iv) assuming the models capture the effects of external forcing (especially concentration changes of greenhouse gases), prediction means, essentially, prediction of long-term (decadal) internal variability. Since both deterministic and stochastic processes contribute to internal variability, ensembles of simulations are required. The effects of initialisation have been discussed by Keenlyside et al. (2008), Pohlmann et al. (2009), Müller et al. (2012), Smith et al. (2012), García-Serrano et al. (2013), and Doblas-Reyes et al. (2013).

For regional scale applications, the information provided by global models is much too coarse, so that regional downscaling to resolutions in the order of 10 km will be necessary. For climate projections and climate assessment, this has been shown to yield added value (Feldmann et al., 2013; Wagner et al., 2013; Berg et al., 2013; Trail et al., 2013). Whether such added value can also be found in regionally downscaled predictions is presently an open question and one of the aims of this study. Another open question is what metrics should be used to measure the skill of predictions, and what metrics are useful for applications. Whereas science is interested in variability and ensemble metrics (Goddard et al., 2013; Gangstøet et al., 2013), practitioners require either categorical (e.g. above/below climatology) or statistical information, such as return values, frequency, and duration of extremes with high spatial resolution (Dool, 2007; Berg et al., 2013). The German Ministry for Education and Research (BMBF) has launched a major programme called MiKlip (Mittelfristige Klimaprognosen, Decadal Climate Prediction, <http://www.fona-miklip.de/en/index.php>) with the aim to establish an operational decadal climate prediction system for Europe, based on the MPI-ESM-LR global model system (Stevens et al., 2013) and the regional climate models (RCMs) COSMO-CLM (Doms and Schättler, 2002; Rockel et al., 2008; Panitz et al., 2013) and REMO (Jacob, 2001) for regional downscaling. The project consists of five modules assessing the different aspects of predictability described above: initialisation, relevant processes, regionalisation, validation, and synthesis. The skill of the predictions will be assessed from an ensemble of decadal hindcasts, which are compared to observations mainly of temperature and precipitation.

This paper describes the regional predictive hindcast ensemble and its validation and discusses the added value of regional downscaling with COSMO-CLM. Section 2 briefly describes the set-up of the MPI-ESM-LR simulations and gives an overview over the set-up of the COSMO-CLM simulations, the construction of the ensemble, and the data used

for validation. Section 3 describes detrending and debiasing, and the validation framework, including the basic set of metrics used. In Sect. 4 we present results and ensemble statistics for Europe. A summary, conclusions, and a brief outlook are given in Sect. 5.

2 Experimental design – construction of the regional decadal ensemble

The aim of MiKlip is to develop a decadal prediction system in several development stages. A first phase has been established with a so called “baseline ensemble” of decadal predictions. It encompasses the global decadal simulations performed with the MPI-ESM (Stevens et al., 2013) according to the CMIP5 protocol (Hurrell et al., 2011). The atmospheric resolution is T63 (1.86°) horizontally with 47 vertical levels up to 0.1 hPa in the vertical. The resolution of the ocean component is 1.5° on average. The ocean is initialised by applying the temperature and salinity anomalies from an ocean-only simulation forced by NCAR/NCEP reanalysis, as described in Matei et al. (2012). A global ensemble is generated from perturbations of the initial atmospheric states with 1-day time lag. The hindcast periods start annually from 1961 to 2012. The ensemble size is 10 members every 5 years (starting dates 1 January 1961, 1966, 1971, and so on) and three members with a starting date 1 January of the in-between years. More details and first results from the global ensemble can be found in Müller et al. (2012). This global baseline ensemble is used as a starting point for the downscaling exercise described here. For our regional ensemble, a larger ensemble for a given starting date was preferred to a higher number of starting dates with a smaller ensemble, since this is a first step to analyse the spread and reliability of the regional ensemble with respect to the global ensemble. The number of starting dates will be increased in the next development stage. On the other hand, a model climatology for the whole period 1961–2010 is necessary to calculate the model anomalies. Therefore, all 10 available realisations of the MPI-ESM-LR 10 year hindcasts for five starting dates (1 January of the years 1961, 1971, 1981, 1991, 2001) were downscaled covering the whole 50 year period. Europe was selected for the regional downscaling.

Two RCMs – namely COSMO-CLM (Consortium for small-scale modelling in climate mode, CCLM hereafter, Doms and Schättler, 2002) and REMO (Jacob, 2001) have been tested. The common simulation domain is chosen according to the CORDEX-EU specifications (Jacob et al., 2013; Giorgi et al., 2009) with a grid resolution of 0.22° and a rotated pole at -162° longitude and 39.25° latitude. The model configuration uses 40 vertical levels. This paper refers only to the simulations with CCLM; the CCLM and REMO simulations will be combined in a later project phase. CCLM is used in the same model version as for CORDEX (cf. Panitz et al., 2013).

All initial and boundary conditions are obtained from the driving model except for the soil. The soil is a compartment of the earth system with long memory and exhibits a considerable spatial heterogeneity, which cannot be fully captured by the global climate model. Therefore, CCLM is initialised using the following strategy. A long-term CCLM reference simulation covering the whole hindcast period is performed forced by reanalysis data. This simulation starts in 1959 using ERA40 (Uppala et al., 2005) as initial and boundary conditions. The first 2 years are used as spin-up. From 1979 until 2010 ERA Interim-forcing (Dee et al., 2011) is applied. At the transition date (1 January 1979), the atmospheric and ocean boundary forcing is switched from ERA40 to ERA Interim, but the soil fields are kept from the ERA40 forced simulation. Additional tests for the first years after the transition from ERA40 to ERA Interim showed only small discrepancies: less than 0.2°C deviation in the European annual mean temperature in the first year (1979), and $< 0.1^{\circ}\text{C}$ after 1980. Therefore, no special treatment for the transition phase has been applied. The initial soil conditions for the CCLM hindcast simulations are then obtained at the respective starting date from this long-term reference simulation. Although there could be some residual drift if the soil moisture climatology under MPI-ESM-LR forcing differs from that under ERA analysis forcing, this approach at least provides a reasonable estimation of the soil moisture initial mean state and anomalies.

To evaluate the model performance, E-OBS v8.0 climatology (Haylock et al., 2008) for near-surface temperature and precipitation was used. E-OBS is a gridded observational data set and available in daily resolution from 1 January 1950 until 31 December 2012. It comprises the variables precipitation, temperature, and sea level pressure in Europe at 25 km over land and is based on ECA&D (European Climate Assessment & Dataset; <http://eca.knmi.nl/>).

3 Data pre-processing and skill metrics

3.1 Data pre-processing

In order to assess the predictive potential of decadal hindcasts in terms of skill (Goddard et al., 2013) and reliability (Corti et al., 2012), the data pre-processing and metrics to be used must be determined. Different approaches can be found in the literature: Bellucci et al. (2013) analyses both anomalies including the long-term trends and anomalies, where the long-term trend has been removed; Goddard et al. (2013) and Müller et al. (2012) use anomalies including the long-term trend, but compare initialised forecasts with uninitialised projections. Following Latif et al. (2010), and partly Bellucci et al., 2013, we decided to extract decadal variability from the time series by removing the long-term means and trends, thus avoiding the problem of interpreting a mixture of long-term and decadal changes.

Another question is about the best practice of removing the long-term trend, which could be non-linear. This problem was assessed by van Oldenborgh et al. (2012), who applied different trend definitions, such as global CO_2 data as the regressor, modelled and observed global temperatures, and simple linear regression. They found that the "... fluctuations in the forecasts and observations are so much larger than the non-linearities in the trend ..." that the "... exact definition of the trend does not affect the results". Finally, they decided to use the global CO_2 data to describe the trend because of physical reasoning. Furthermore, van Oldenborgh et al. (2012) used the global data to detrend on grid point basis and stated that the global trend also described a large part of the data on the regional scale. In contrast, due to small trends and large variability, van Oldenborgh et al. (2012) did not remove the trends for precipitation and rather analysed the hindcast skill of the un-modified data. However, we are explicitly interested in regional differences and the different seasons (summer/winter). Based on climate projections, the Intergovernmental Panel on Climate Change (IPCC; Christensen et al., 2007) found: (i) annual temperatures in Europe are likely to increase more than the global mean, (ii) northern Europe will experience a larger warming in winter and southern Europe in summer, and (iii) annual precipitation will be increased in northern Europe, whereas a decrease in southern Europe will be likely observed. Due to the fact that the response to greenhouse gas forcing depends on the region and season, we decided to remove the long-term trend on grid point basis using simple linear regression. This procedure additionally assures zero-mean residuals on grid basis. We applied the regression to both temperature and precipitation. More sophisticated approaches of de-trending exist, e.g. the Empirical Mode Decomposition (EMD) method (Wu et al., 2007), and could be a valuable alternative.

To illustrate the problems associated with decadal variability, an E-OBS anomaly time series of summer half-year precipitation sums at a grid point in Germany is shown in Fig. 1.

The detrended and unfiltered anomalies are shown as a thin line. The summer to summer variability (thin line) is high, and these high frequency fluctuations are unlikely to be predictable using decadal model initialisations. On the other hand, the low pass filtered data (thick line) appear more likely to be predictable by decadal predictions.

Spatial smoothing of the data is beneficial in skill assessment due to reduction of grid-scale noise (Räisänen and Ylhäisi, 2011). Goddard et al. (2013) advocate a 5° latitude \times 5° longitude spatial smoothing for precipitation and a 10° latitude \times 10° longitude smoothing for temperature, which is not appropriate for the regionalisation purpose. They also present analyses on different time scales, i.e. year 1, years 2–5, years 6–9 and years 2–9, to discuss the effect of different lead times and temporal averaging. In contrast to Goddard et al. (2013), whose data sets start annually, we are limited to five starting dates in the first regional ensemble generation, as explained above, and, therefore, have a

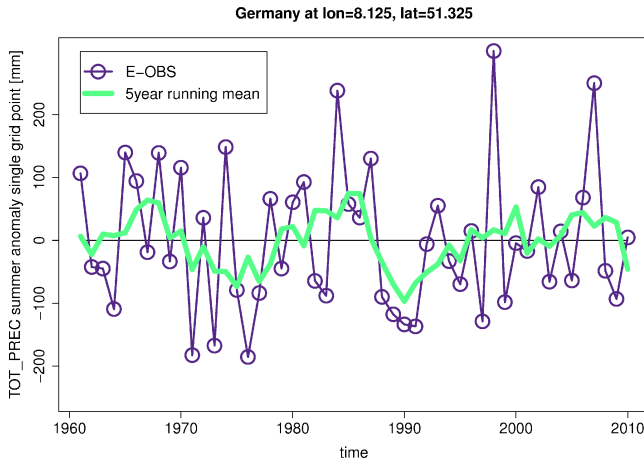


Figure 1. Summer half-year observational (E-OBS) precipitation anomalies at a single grid point in western Germany (circles connected with a thin line). The long term mean and trend have been removed by a linear regression. A moving average filter of 5 years is applied to illustrate variability on multi-year time scales (green line).

much smaller data sample available. As a compromise, we decided to split the data into two parts, one with lead times 1–5 years and the other with lead times 6–10 years. The first data set can be considered as representing the skill which mainly originates from the initialisation, while the second data set is more related to the representation of slow varying climate components. Note that for each prediction horizon, lead times are not averaged, as suggested in Goddard et al. (2013), but are considered together successively, yielding 25 data points for each part of the time series.

Since we are interested in the regional scale, we will analyse the data at their original 25 km resolution without spatial smoothing. For comparison of the global model with the regional model, we will interpolate the global MPI-ESM-LR data to the 25 km grid.

To summarise, our pre-processing consists of (i) aggregating half-year precipitation sums and temperature averages, (ii) removing long-term means and trends, and (iii) splitting the data into two parts: the first with lead times 1–5 years and the second with lead times 6–10 years.

3.2 Metrics

The following metrics will be used to characterise the CCLM and MPI-ESM-LR ensembles vs. observations and to identify the potential added value:

- Skill:

To quantify the predictive skill of the CCLM and MPI-ESM-LR ensembles against observations we will use the Pearson correlation coefficient ρ applied to anomalies (also known as ACC (anomaly correlation coefficient); Bellucci et al., 2013). If we denote the anomaly

ensemble mean at a specific location i as $m_{t,i}$ and the corresponding observed anomalies as $o_{t,i}$, where $t = 1, \dots, N$ represents the time index with $N = 25$ data points (semi-annual means 1961–2010, 5 lead years), the correlation coefficient is given by

$$\rho_i = \frac{1/N \cdot \sum_t m_{t,i} \cdot o_{t,i}}{\sigma_{m_i} \cdot \sigma_{o_i}}. \quad (1)$$

We have estimated the statistical significance of the correlation coefficient on grid point basis with the test statistic:

$$t = \frac{\rho}{\sqrt{(1 - \rho^2)/N_{\text{eff}}}} \quad (2)$$

on a t distribution with N_{eff} (two-sided) degrees of freedom. Due to serial autocorrelations the effective number of degrees of freedom is reduced. Therefore, we account for autocorrelations according to Wilks (2011):

$$N_{\text{eff}} = N \cdot \frac{1 - \phi}{1 + \phi}, \quad (3)$$

where ϕ is the autocorrelation at lag 1 of $[m_{t,i} \cdot o_{t,i}]$. Due to the temporal gaps in the data we used the Discrete Autocorrelation Function developed by Edelson and Krolik (1988). Although only lag 1 is considered explicitly, further autocorrelations at higher lags are also considered implicitly, since the correction approach is based on autoregressive processes of order 1 (AR[1]) (Thiebaut and Zwiers, 1984; Mieruch et al., 2014), whose autocorrelation function decreases exponentially, thus considering also higher lags than 1. For uncorrelated data, the statistical significance on the 10% level for 25 data points is achieved by a correlation coefficient of $|\rho| \geq 0.33$. Only in few cases we will observe correlations fulfilling the significance criterion, taking into account serial autocorrelations (indicated as stippling in the respective figures). Mostly we observe “non-significant” correlations between the model data and the observations. Such results are difficult to interpret in the sense of statistical hypothesis testing. For instance, von Storch and Zwiers (2013) claim that “... a statistical null hypothesis may not be a well-posed problem ...” and “Even if statistical testing were completely appropriate, the dependency of the power of statistical tests on the sample size n remains a limitation on interpretation”. We therefore follow von Storch and Zwiers (2013) who proposed “... a simple descriptive approach for characterising the information in an ensemble ...”. This means a hypothesis test will be performed, but we will not completely rely on the significance, especially because we are dealing with small sample sizes (25) on grid pixel basis and the power of the test is questionable. Thus, if we observe weak positive correlations for Europe between model and observations in 90% of the

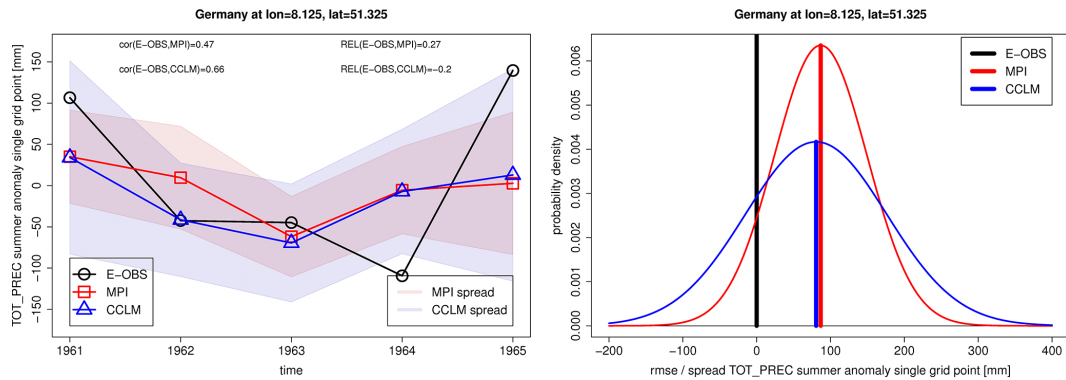


Figure 2. Snapshot from our time series showing the “added value” of downscaling, based on real data.

grid pixels, we believe that there is a certain relationship, even if it is not significant on a single grid pixel basis.

– Reliability:

Concerning the reliability of an ensemble forecast we follow Weigel et al. (2009) who define reliability as a measure of “how consistent the forecast probabilities are with the relative frequencies of the observed outcomes” (cf. also Mason, 2008) and give the following definition:

$$REL_i = \frac{RMSE(\mu_{t,i}, x_{t,i}) - \sqrt{\langle \sigma_{i,ens}^2 \rangle_t}}{RMSE(\mu_{t,i}, x_{t,i})}, \quad (4)$$

where the index i indicates a single grid point and t is the time index. According to this definition, and to interpret the results correctly, Weigel et al. (2009) explain that a normally distributed ensemble is reliable if, and only if, the root mean square error (RMSE) between the ensemble mean and the observations is identical to the time-mean ensemble spread $\sqrt{\langle \sigma_{ens}^2 \rangle}$. Temperatures easily fulfil the normality assumption and even half-year precipitation sums are quite normally distributed, due to the central limit theorem. The ensemble is called underconfident ($REL < 0$) if $RMSE(\mu, x) < \sqrt{\langle \sigma_{ens}^2 \rangle}$; it is called overconfident ($REL > 0$) if $RMSE(\mu, x) > \sqrt{\langle \sigma_{ens}^2 \rangle}$, and calibrated ($REL = 0$) if $RMSE(\mu, x) = \sqrt{\langle \sigma_{ens}^2 \rangle}$. Loosely speaking, reliability measures if the ensemble spread covers the model errors. Underconfidence is generally considered less harmful than overconfidence, as long as the forecasts/hindcasts have similar predictive skill. To test the reliability on statistical significance we used a two-sided F-test with the null hypothesis $H_0 : MSE(\mu, x) = \langle \sigma_{ens}^2 \rangle$, that the mean square error is similar to the ensemble spread. The test statistic

is given by

$$F = \frac{MSE(\mu, x)}{\langle \sigma_{ens}^2 \rangle} \quad \text{if } MSE(\mu, x) > \langle \sigma_{ens}^2 \rangle \quad \text{and} \quad (5)$$

$$F = \frac{\langle \sigma_{ens}^2 \rangle}{MSE(\mu, x)} \quad \text{if } MSE(\mu, x) < \langle \sigma_{ens}^2 \rangle \quad (6)$$

and evaluated on an F distribution with $N_{eff}^{MSE} = N_{eff}^{\sigma}$ degrees of freedom (based on the original $(N - 1) = 24$ data points) according to Eq. (3). The autocorrelations at lag 1 are estimated using the Discrete Autocorrelation Function on the time series (for each grid pixel) of the $(\mu_{t,i} - x_{t,i})^2$ and $\sigma_{ens,t}^2$. It turned out that the null hypothesis cannot be rejected on the 10 % level as long as REL lies approximately in the range of ± 0.2 . Cases where we have to accept the null hypothesis, i.e. the model is “reliable”, will be indicated by stippling in the respective figures.

4 Decadal variability assessment

4.1 An example

We illustrate the idea of added value using a typical time series from 1961 to 1965 at a grid point (8.125° longitude and 51.325° latitude) in Germany shown in the left panel of Fig. 2.

The black time series in Fig. 2 are the E-OBS summer precipitation sums, the red data are the MPI-ESM-LR simulations and the blue data are the CCLM hindcasts. Large-scale decadal predictability is inherited from the global model to the regional one: if there is no predictability in the global model, there will be no predictability in the regional model. We expect that due to the higher resolution and better representation of small-scale processes, the regional model will be closer to the observations. The left panel of Fig. 2 shows that the regional model uses the already skillful global simulations to move a bit closer to the observation, which results

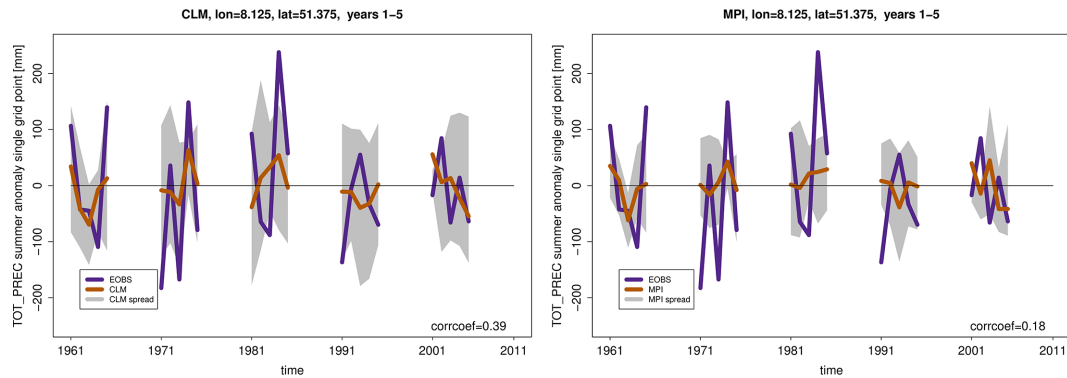


Figure 3. Left: Summer half-year precipitation sums of E-OBS (lilac) and CCLM ensemble mean (orange) at a location in Germany for lead years 1–5. The correlation coefficient is 0.39. The grey shaded area depicts the CCLM ensemble spread, i.e. the standard deviation over the ensemble. Right: Same location but for MPI-ESM-LR ensemble mean (orange). The correlation coefficient is 0.18.

in a slightly better correlation between the regional model and observations (0.66) than for the global model (0.47). The ensemble spread of the global model is too small and is indicated by the reliability of 0.27. Due to the downscaling, the spread is increased (reliability is -0.2) and thus accounts better for the uncertainties. Additionally, the concept of reliability is schematically shown in the right panel of Fig. 2. The black line depicts the E-OBS observations and the MPI-ESM-LR (red) and CCLM (blue) data are represented by Gaussians, with the respective RMSE as mean and the ensemble spread as standard deviation. Again, it can be seen that due to the downscaling, the hindcast moves slightly closer to the observation and simultaneously the spread is increased, yielding a higher probability for the E-OBS outcome.

Thus, the “added value” which we expect from the downscaling is an increase of skill together with an improvement of the ensemble spread. This requirement is not trivial: ensemble recalibration techniques, such as the CCR (Climate Conserving Recalibration, Weigel et al., 2009), are able to increase the ensemble spread, but at the cost of a loss in correlation.

Clear signals of such an “added value” of the downscaling can be observed for summer precipitation for lead years 1–5, as shown in Fig. 3.

The left panel of Fig. 3 shows yearly E-OBS data in lilac and the CCLM ensemble mean anomalies in orange. Additionally, the CCLM ensemble spread, i.e. the standard deviation over the ensemble for each time step, is shown as the grey shaded area. Due to the decadal initialisation of CCLM in 1961, 1971, 1981, 1991, and 2001, we have separated the decades from each other and show the lead years 1–5. The correlation coefficient is 0.39. The right panel shows MPI-ESM-LR data at the same location. The correlation between MPI-ESM-LR and E-OBS is 0.18. We observe a clear improvement of the correlation using CCLM, which arises from moving closer to the observations and, hence, also better de-

scription of the low frequency variability. Concerning the reliability, it can clearly be seen that the CCLM spread (left) is much larger than the MPI-ESM-LR spread (right) and covers the observations in most cases. This indicates that single CCLM ensemble members show very similar variability to the observations, whereas single MPI-ESM-LR ensemble members are overconfident. It appears that, due to the regionalisation, we are able to increase the predictive skill, i.e. the correlation and simultaneously the ensemble spread. However, it is worth to recall that when there is no skill in MPI-ESM-LR, there is no skill in CCLM.

Another aspect relates to the long-term performance of the models. The long-term means and trends of our regional simulations belong (by definition) not to the quantities, which vary on decadal time scales; hence, they are not the subject of our assessment of decadal predictability. The long-term performance of the initialised predictions is, in principle, similar to the long-term performance of the projections (Feldmann et al., 2013; Wagner et al., 2013; Berg et al., 2013; Trail et al., 2013).

4.2 Summer precipitation

4.2.1 Skill and reliability: years 1–5

Figure 4 shows the correlation coefficient and the reliability for Europe of summer precipitation sums for lead years 1–5. The top left panel of Fig. 4 presents the correlation between CCLM and E-OBS and the top right panel shows the correlations between MPI-ESM-LR and E-OBS. The bottom panels of Fig. 4 display the respective reliabilities. The stippling indicates results, significant at the 10 % level.

To judge the decadal predictability, both correlation and reliability should be evaluated together. In large parts of Europe we observe an increase in correlation between the regional CCLM model and E-OBS, with respect to the MPI-ESM-LR. This comprises the British Isles, the Benelux region, the northern part of France, Germany, Poland, the

Summer precipitation anomalies, lead years 1-5

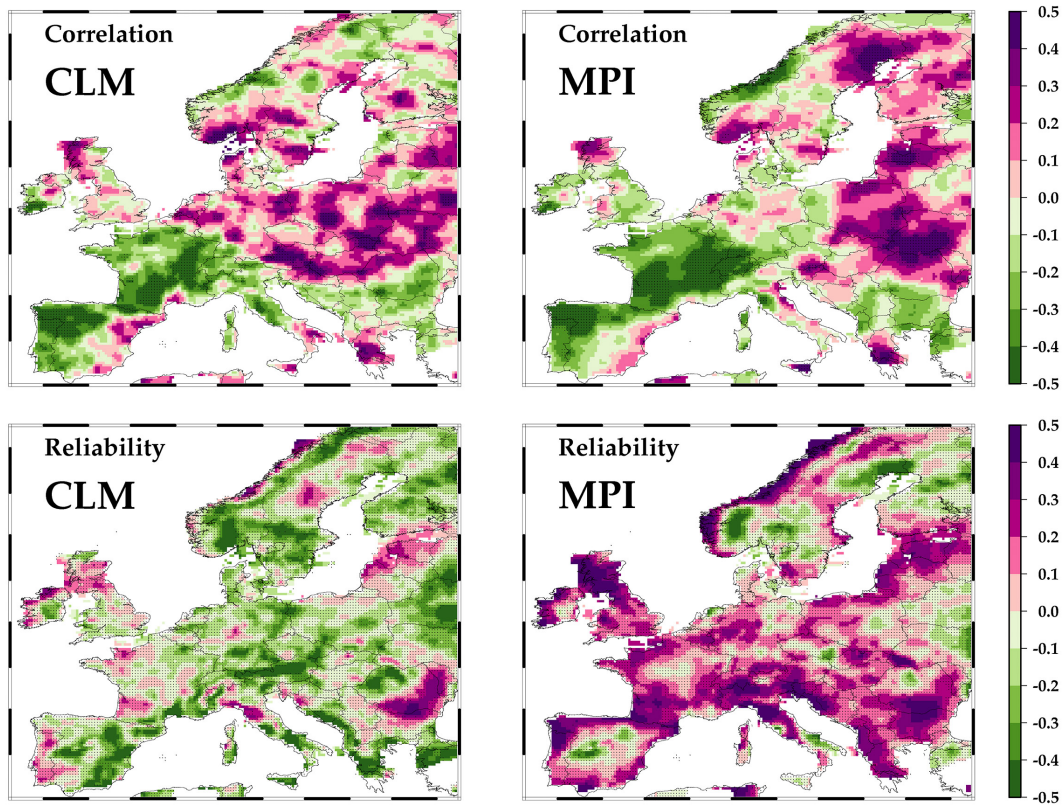


Figure 4. Summer half-year precipitation anomaly sums for lead years 1–5 from 1961 to 2010. The top panels show the correlation coefficient between E-OBS observations and CCLM (left) and MPI-ESM-LR (right). The bottom panels show the reliability of the CCLM (left) and MPI-ESM-LR (right) ensembles with respect to the E-OBS observations. Stippling indicates results, significant at the 10 % level.

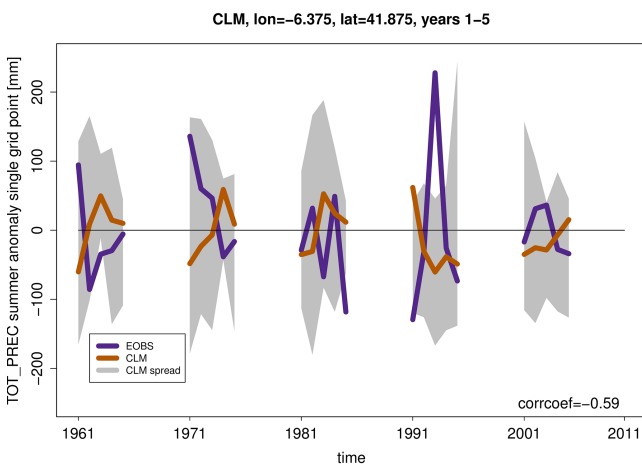


Figure 5. CCLM summer precipitation for lead years 1–5 in northern Spain at the border to Portugal.

Czech Republic, and Austria, as well as Scandinavia. There seems to be a small loss in skill over France, the Mediterranean region and more or less skill preservation in eastern Europe. However, there are also regions with negative

skill, e.g. in Portugal and northern Spain. A negative anomaly correlation at the Iberian Peninsula has also been found by Bellucci et al. (2013) for a multi-model ensemble. Figure 5 shows a time series in northern Spain at the border to Portugal.

Here, a correlation of -0.59 has been computed and it is clearly seen how the time series evolve in different directions. The quite strong correlation indicates that opposed dynamics are probably not by chance, but rather they are systematic. The reason for such behaviour cannot be explained within the scope of this study.

In summary, we can say that downscaling is, in general, beneficial and the CCLM can add value to the global driving data.

Regarding reliability, shown in the bottom panels of Fig. 4, the regional model clearly improves the results. Based on a significance test on a grid basis, it is evident that good values of the reliability for the CCLM and MPI-ESM-LR hindcasts lie approximately within ± 0.2 , which is indicated by stippling in the respective figures. However, a reliability of -0.2 is preferable to a reliability of $+0.2$ (as long as both have the same skill), since underconfident (negative REL)

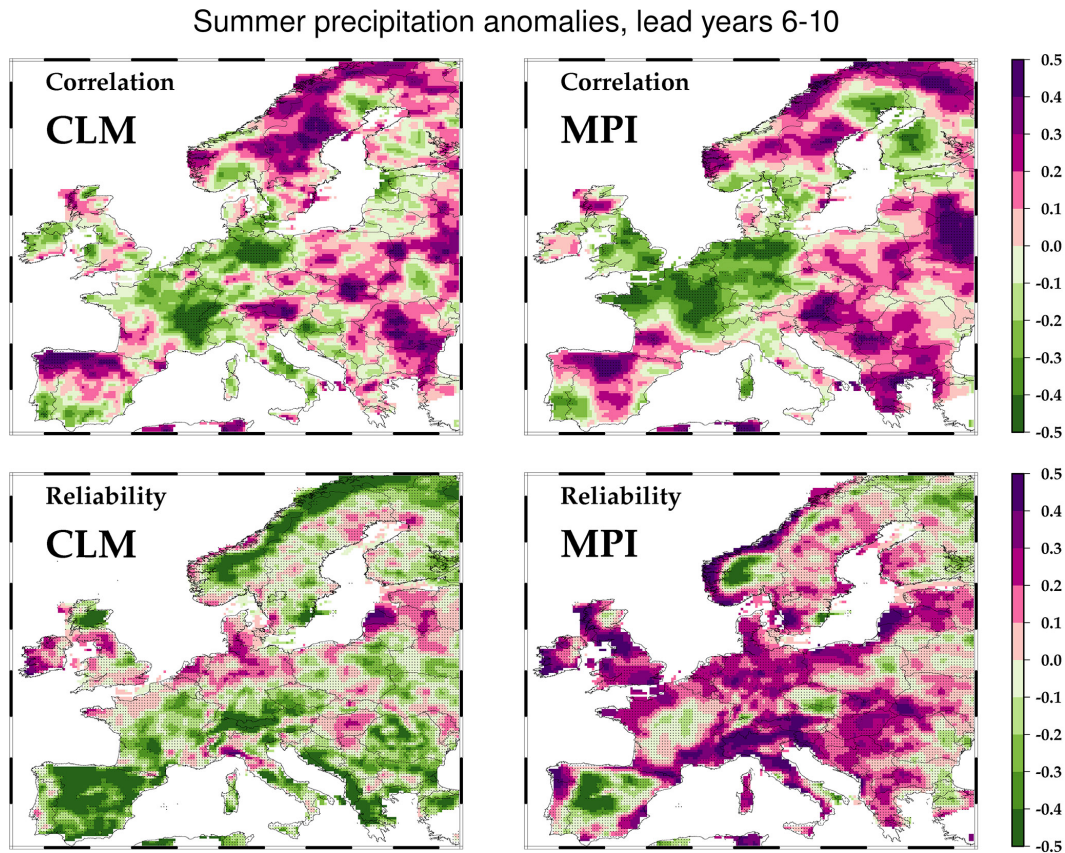


Figure 6. Summer half-year precipitation anomaly sums for lead years 6–10 from 1961 to 2010. The top panels show the correlation coefficient between E-OBS observations and CCLM (left) and MPI-ESM-LR (right). The bottom panels show the reliability of the CCLM (left) and MPI-ESM-LR (right) ensembles with respect to the E-OBS observations. Stippling indicates results, significant at the 10 % level.

ensembles better account for uncertainties, and overconfident ensembles (positive REL) tend toward incorrect predictions. The MPI-ESM-LR (bottom right panel of Fig. 4) shows too much overconfidence and cannot reproduce the uncertainties reliably. Thus we can state an added value of the regionalisation in the sense explained in Sect. 4.1, namely the increase of predictive skill and simultaneously improving the ensemble spread.

4.2.2 Skill and reliability: years 6–10

Figure 6 shows the correlation coefficient and the reliability for Europe of summer precipitation sums for lead years 6–10. The top left panel of Fig. 6 presents the correlation between CCLM and E-OBS and the top right panel shows the correlations between MPI-ESM-LR and E-OBS. The bottom panels of Fig. 6 display the respective reliabilities. The stippling indicates results, significant at the 10 % level.

The decadal predictions of lead years 6–10 are not so strongly influenced by the initialisation at the beginning of the decades. Of higher importance is if the model was able to capture low frequency modes such as NAO or AMO. Interesting cases of positive and negative correlations are found,

where the CCLM in general shows slightly higher correlations, at least the preservation of skill. In eastern Europe for example we have skill observed in lead years 1–5 and also in years 6–10. Thus it seems that the initialisation at the beginning of the decade is beneficial and also longer term climate signals could be reproduced. In Central Europe, a good skill in lead years 1–5 turns into a relative strong anti-correlation in lead years 6–10. Here we speculate that the initialisation is advantageous, but after about 5 years the model (MPI-ESM-LR) undergoes a kind of phase shift. Contrastingly, in South Europe (especially Iberian Peninsula) the lead years 1–5 showed negative correlations and for lead years 6–10 we have observed a positive skill. Thus, it seems that the model initially moves into the wrong (phase shifted) direction and after about 5 years it is able to swing into the correct phase of low frequency climate signals. As mentioned above, the reason for such a behaviour cannot be explained within this study. One possible cause could be explained by a beat, whereas the model exhibits a slightly different frequency than the real frequency of e.g. NAO or AMO.

The reliability, shown in Fig. 6 (bottom panels) is much improved by the downscaling and together with the skill im-

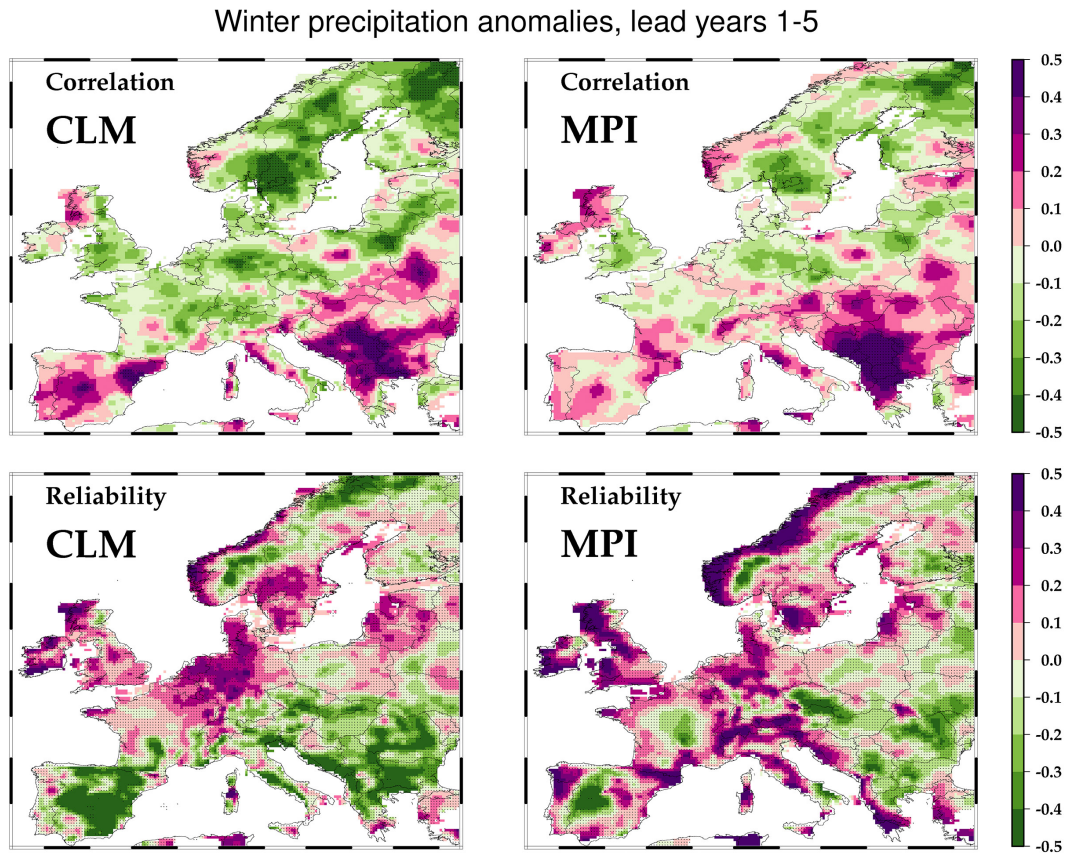


Figure 7. Winter half-year precipitation anomaly sums for lead years 1–5 from 1961 to 2010. The top panels show the correlation coefficient between E-OBS observations and CCLM (left) and MPI-ESM-LR (right). The bottom panels show the reliability of the CCLM (left) and MPI-ESM-LR (right) ensembles with respect to the E-OBS observations. Stippling indicates results, significant at the 10 % level.

provement (preservation) constitutes again the added value of dynamical downscaling global decadal predictions.

4.3 Winter precipitation

4.3.1 Skill and reliability: years 1-5

Figure 7 shows the correlation coefficient and the reliability for Europe of winter precipitation sums for lead years 1–5. The top left panel of Fig. 7 presents the correlation between CCLM and E-OBS and the top right panel shows the correlations between MPI-ESM-LR and E-OBS. The bottom panels of Fig. 7 display the respective reliabilities. The stippling indicates results, significant at the 10 % level.

Positive predictive skill is in principle only found in southern Europe, i.e. Iberian Peninsula, Italy and South-East Europe. Similar to the results from the analysis of summer precipitation, downscaling slightly increases the correlation, e.g. at the Iberian Peninsula. This amplification process works also for negative correlations. For instance in central Germany, the MPI-ESM-LR yields weak negative correlations. The regional model amplifies the correlations and achieves even the 10 % significance level. As discussed

above the reason for strong negative correlations is up to now unclear.

The reliability of the regional model is mainly improved compared to the global model, but slightly to underconfident in parts of southern Europe.

4.3.2 Skill and reliability: years 6–10

Figure 8 shows the correlation coefficient and the reliability for Europe of winter precipitation sums for lead years 6–10. The top left panel of Fig. 8 presents the correlation between CCLM and E-OBS and the top right panel shows the correlations between MPI-ESM-LR and E-OBS. The bottom panels of Fig. 8 display the respective reliabilities. The stippling indicates results, significant at the 10 % level.

The results for winter precipitation for lead years 6–10 are very similar to the finding at the beginning of the decade. However, the correlations are slightly weaker in the second half of the decade and the reliability is comparable. Again downscaling improves the reliability especially in southern Europe.

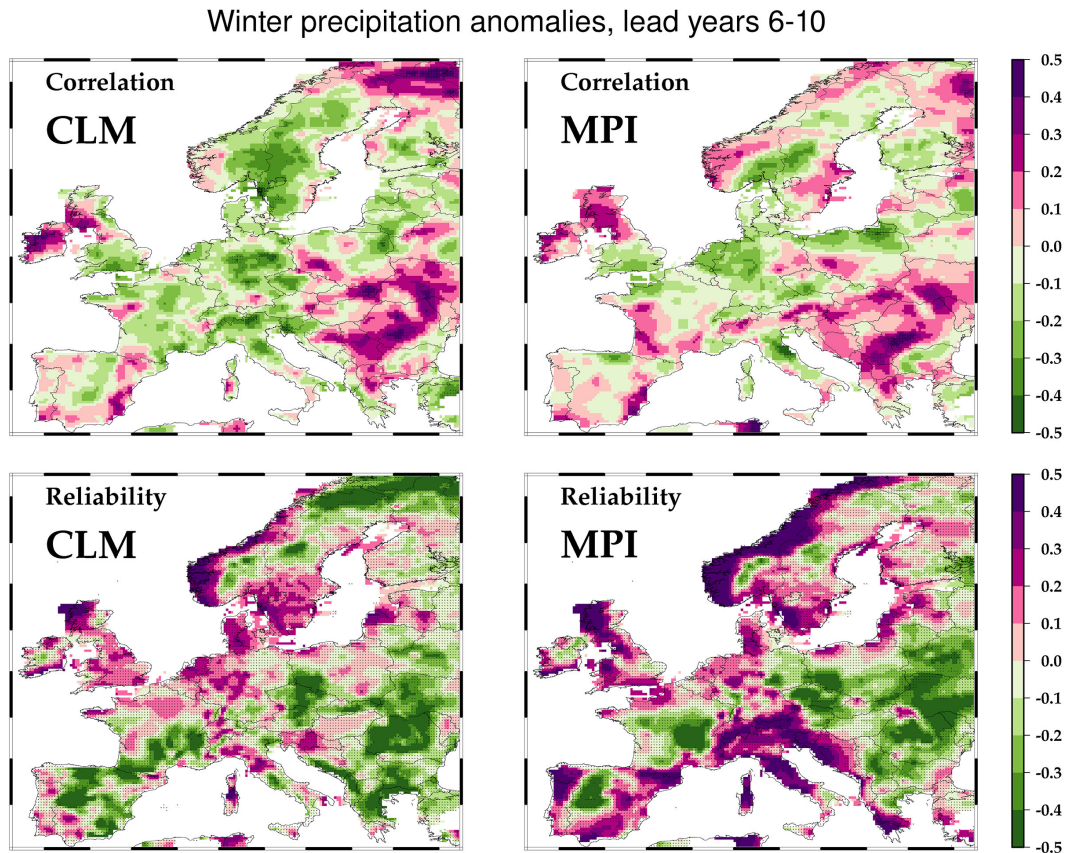


Figure 8. Winter half-year precipitation anomaly sums for lead years 6–10 from 1961 to 2010. The top panels show the correlation coefficient between E-OBS observations and CCLM (left) and MPI-ESM-LR (right). The bottom panels show the reliability of the CCLM (left) and MPI-ESM-LR (right) ensembles with respect to the E-OBS observations. Stippling indicates results, significant at the 10 % level.

4.4 Summer temperatures

4.4.1 Skill and reliability: years 1-5

Figure 9 shows the correlation coefficient and the reliability for Europe of summer temperature anomaly means for lead years 1–5. The top left panel of Fig. 9 presents the correlation between CCLM and E-OBS and the top right panel shows the correlations between MPI-ESM-LR and E-OBS. The bottom panels of Fig. 9 display the respective reliabilities. The stippling indicates results, significant at the 10 % level.

Most regions of Europe show similar positive correlations in both the MPI-ESM-LR and CCLM models. The correlations are mostly non-significant at the 10 % level on grid pixel basis. However, according to von Storch and Zwiers (2013), hypothesis testing on model ensembles has to be interpreted carefully. The standard interpretation of non-significant correlations would be that the observed correlations are more or less found by chance. Inspecting the positive correlations covering almost all of Europe (top panels of Fig. 9), it is hard to believe that such patterns are a fortunate coincidence. Thus, in spite of the weak correlations, the small sample population on grid basis, as well as serial and

spatial autocorrelations, there seems to be some decadal predictive skill in the model. Although evidence for this statement is not provided, a look into a typical time series may yield more confidence. Figure 10 shows CCLM summer temperature anomalies from a grid point near Exeter in southern England. The left panel shows the lead times of 1–5 years and typical weak correlations are found.

The right panel depicts only the lead times of years 1–2, where it can be seen that in 4 out of 5 decades (1961, 1971, 1991, 2001), the CCLM model evolves in the correct direction during the first 2 years. It seems plausible that we observe higher skill at the very beginning of the decade, shortly after the initialisation. Despite non-significant results, according to the hypothesis test, there appears to be predictive skill in the models. However, due to the small sample size, a definitive conclusion cannot be made.

Regarding the potential added value of the regional model, downscaling appears not to be beneficial for summer temperature anomalies. This is likely due to half-year temperature anomalies that cannot be attributed to a small-scale process and, hence, regionalisation only preserves the skill, and does not improve it. A half-year temperature anomaly in a

Summer temperature anomalies, lead years 1-5

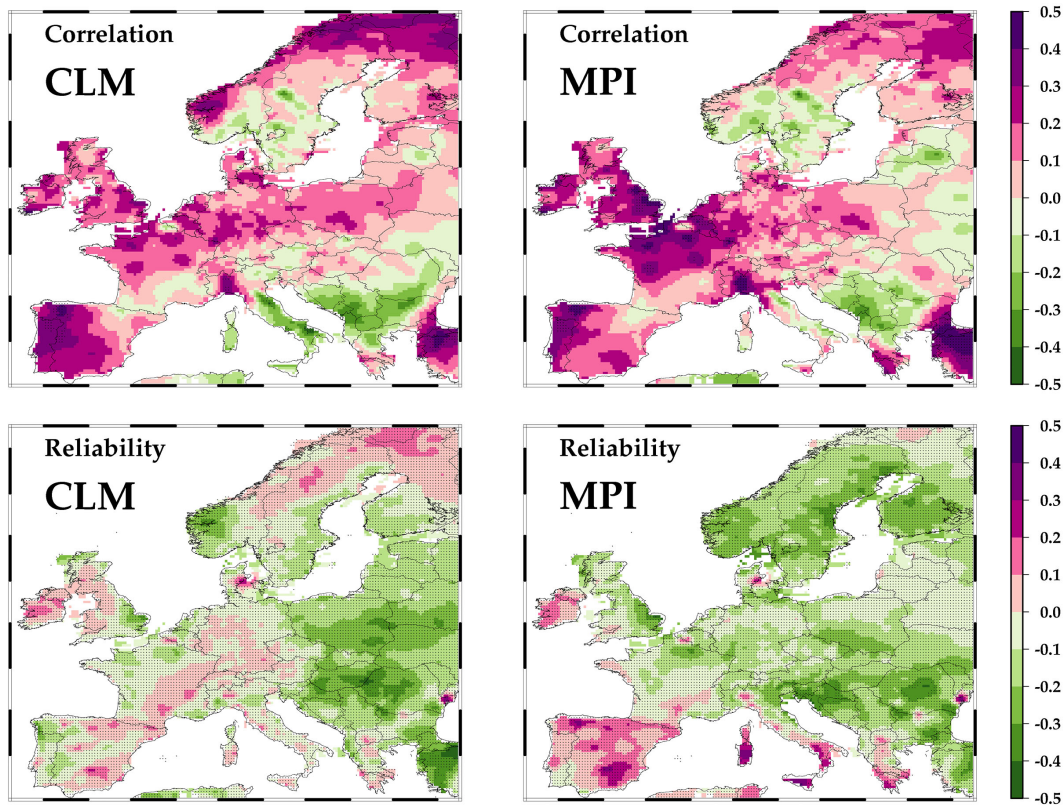


Figure 9. Summer half-year temperature anomaly means for lead years 1–5 from 1961 to 2010. The top panels show the correlation coefficient between E-OBS observations and CCLM (left) and MPI-ESM-LR (right). The bottom panels show the reliability of the CCLM (left) and MPI-ESM-LR (right) ensembles with respect to the E-OBS observations. Stippling indicates results, significant at the 10 % level.

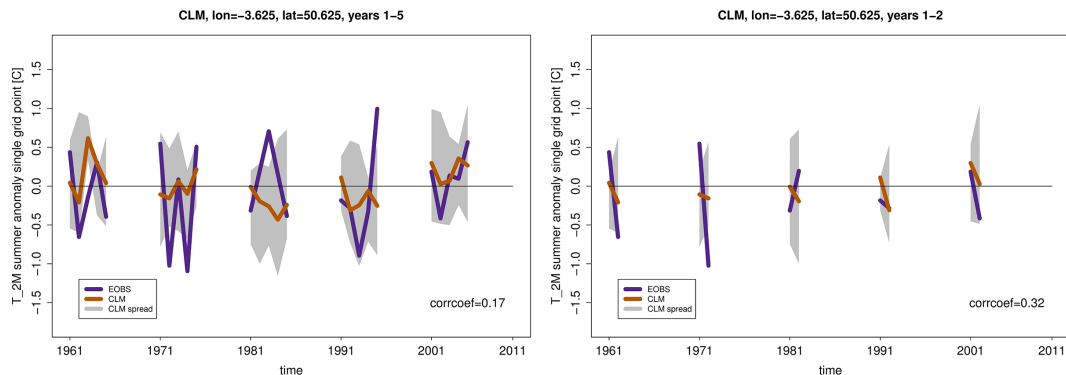


Figure 10. Summer temperatures at a grid point near Exeter in South England. Left: Lead years 1–5. Right: Lead years 1–2.

200 × 200 km box does not differ distinctively to a temperature anomaly in a 25 × 25 km box.

In addition, the reliability patterns (bottom panels of Fig. 9) show similarities between MPI-ESM-LR and CCLM with values within ±0.2. This is also indicated by the stippling denoting no significant difference between the MSE(model, observations) and the model spread on the 10 % level.

4.4.2 Skill and reliability: years 6–10

Figure 11 shows the correlation coefficient and the reliability for Europe of summer temperature anomalies for lead years 6–10. The top left panel of Fig. 11 presents the correlation between CCLM and E-OBS and the top right panel shows the correlations between MPI-ESM-LR and E-OBS. The bot-

Summer temperature anomalies, lead years 6-10

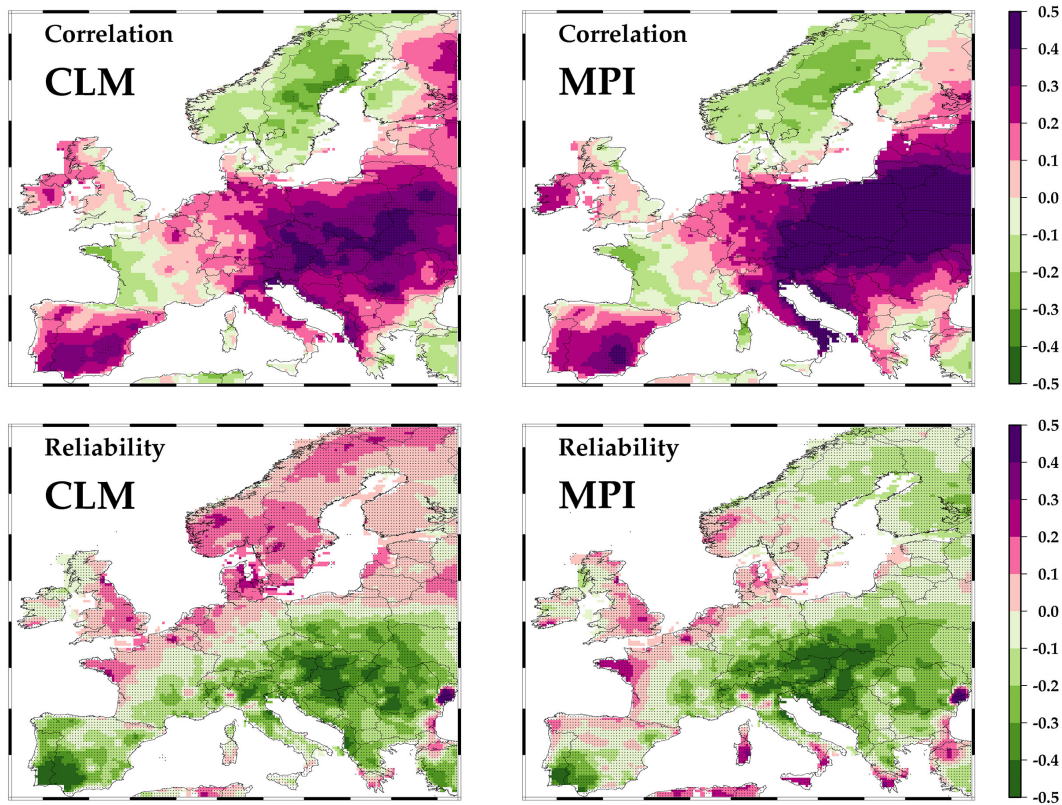


Figure 11. Summer half-year temperature anomaly means for lead years 6–10 from 1961 to 2010. The top panels show the correlation coefficient between E-OBS observations and CCLM (left) and MPI-ESM-LR (right). The bottom panels show the reliability of the CCLM (left) and MPI-ESM-LR (right) ensembles with respect to the E-OBS observations. Stippling indicates results, significant at the 10 % level.

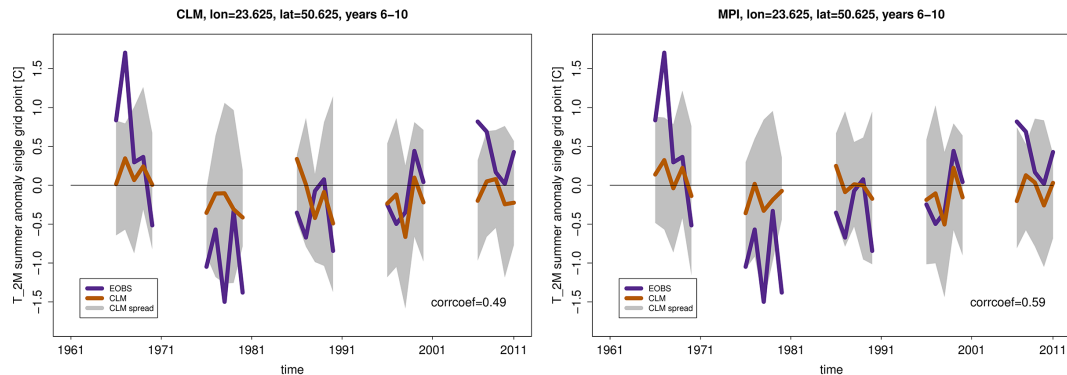


Figure 12. Summer temperature anomalies in South-East Poland. Shown are lead years 6–10.

tom panels of Fig. 11 display the respective reliabilities. The stippling indicates results, significant at the 10 % level.

As can be seen, there is a large similarity between the correlations of the global and the regional model, thus the predictive skill could be preserved but not improved. A strong significant (on grid pixel basis) skill is observed in eastern Europe. Figure 12 shows time series from South-East Poland. It can be seen that there is a left residual trend in the E-

OBS anomalies (lilac). The period 1961–1970 was recorded with higher temperatures than normal, and therefore potentially influence the linear trend estimations strongly during the 1961–2010 time period.

However, this residual trend accounts only for a minor part of the observed predictive skill. This is demonstrated well in the right panel of Fig. 12, which shows the MPI-ESM-LR results. The residual trend is minimal, but in most decades

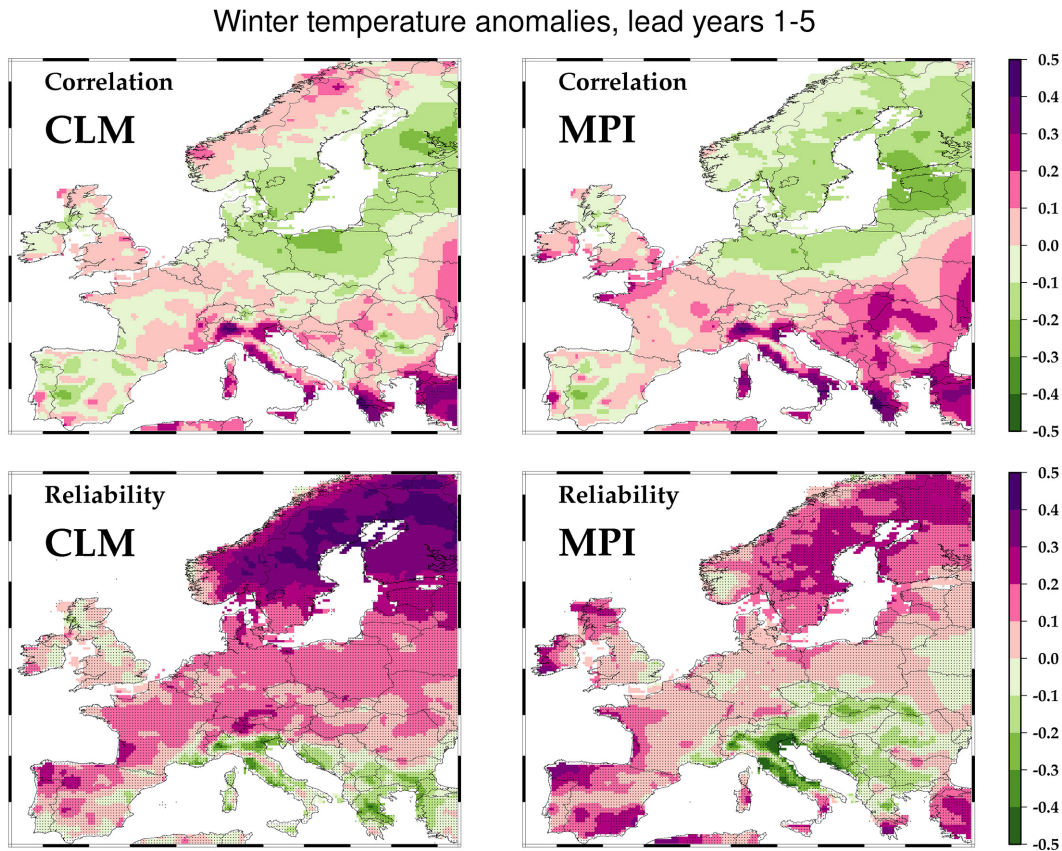


Figure 13. Winter half-year temperature anomaly means for lead years 1–5 from 1961 to 2010. The top panels show the correlation coefficient between E-OBS observations and CCLM (left) and MPI-ESM-LR (right). The bottom panels show the reliability of the CCLM (left) and MPI-ESM-LR (right) ensembles with respect to the E-OBS observations. Stippling indicates results, significant at the 10 % level.

the year-to-year variations are reproduced with weaker amplitude due to the ensemble mean.

The reliabilities of the MPI-ESM-LR and the CCLM are comparable in a satisfactory region within ± 0.2 .

4.5 Winter temperatures

4.5.1 Skill and reliability: years 1–5

Figure 13 shows the correlation coefficient and the reliability for Europe of winter temperature anomaly means for lead years 1–5. The top left panel of Fig. 13 presents the correlation between CCLM and E-OBS and the top right panel shows the correlations between MPI-ESM-LR and E-OBS. The bottom panels of Fig. 13 display the respective reliabilities. The stippling indicates results, significant at the 10 % level.

Again, the correlation pattern for winter temperature anomalies of MPI-ESM-LR and CCLM are very similar. Nevertheless, we found regions where the skill is slightly larger for CCLM, e.g. at the northern coast of Norway and where it is slightly smaller, e.g. in South-East Europe. In general, we have mostly a preservation but no improvement

of skill. The correlations are generally weak, but stronger in the south in contrast to the summer temperatures, where the correlations are larger in the north. The reliability is also comparable between MPI-ESM-LR and CCLM, but it appears downscaling worsens the reliability slightly. The reason is not a smaller spread, but a slightly larger RMSE, which yields slightly more overconfident results.

4.5.2 Skill and reliability: years 6–10

Figure 14 shows the correlation coefficient and the reliability for Europe of winter temperature anomalies for lead years 6–10. The top left panel of Fig. 14 presents the correlation between CCLM and E-OBS and the top right panel shows the correlations between MPI-ESM-LR and E-OBS. The bottom panels of Fig. 14 display the respective reliabilities. The stippling indicates results, significant at the 10 % level.

The correlations of winter temperatures of lead years 6–10 are stronger than for lead years 1–5, which seems to be counterintuitive. A possible cause could be, as speculated above, the global model generates a slightly different low frequency variability than the observations, resulting in being slightly out of phase at the beginning of the decade and later within

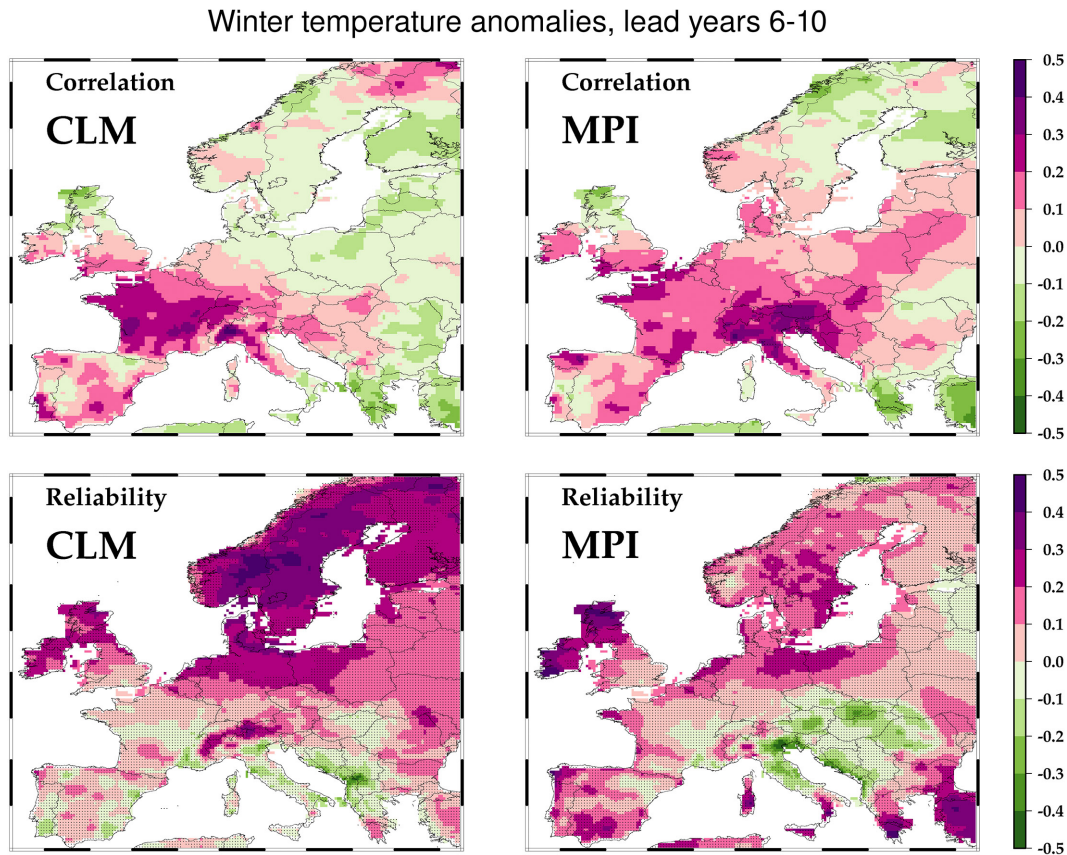


Figure 14. Winter half-year temperature anomaly means for lead years 6–10 from 1961 to 2010. The top panels show the correlation coefficient between E-OBS observations and CCLM (left) and MPI-ESM-LR (right). The bottom panels show the reliability of the CCLM (left) and MPI-ESM-LR (right) ensembles with respect to the E-OBS observations. Stippling indicates results, significant at the 10 % level.

phase at the end of the decade. However, definitive conclusions are not within the scope of this study.

With regards to reliability, both models are slightly overconfident. A small loss in correlation of the CCLM yields to a slightly larger RMSE for CCLM and to a small increase of the values of the reliability.

5 Summary and conclusions

In this study we have analysed regional climate predictions downscaled from global decadal hindcasts using a regional model. We generated a 10 member ensemble of climate simulations (1961–2010) with the regional model CCLM at 25 km resolution, driven by decadal predictions of the global model MPI-ESM-LR. These model runs have been initialised every 10 years from 1961 to 2001. Decadal variability of detrended anomalies of summer and winter precipitation, as well as of summer and winter temperatures, in Europe were compared with observations and results of the global model in order to identify a possible added value of regionalisation. We define the added value of the regional model as an increase of predictive skill and a simultaneous improvement of

the reliability. To quantify the predictability we used the correlation coefficient to measure the skill and the RMSE and ensemble spread to characterise reliability (cf. Sect. 3.2).

Predictive skill, covering almost all of Europe, has been found for summer temperature anomalies at lead years 1–5. This skill has been induced by the global MPI-ESM-LR model and could be preserved, but mostly not improved, by regionalisation. This could be due to the well-known fact that spatial patterns of mean temperature anomalies are quite homogeneous and, therefore, are already well captured at the resolution of the global model so there is little room for improvement due to downscaling. The reliability of summer temperature anomalies is good for both models, which is a typical feature of an ensemble with a low signal-to-noise ratio. A possible option to increase the skill could be an increase in the ensemble size (Scaife et al., 2014a).

This situation is similar for lead years 6–10. Here, a good skill is observed in eastern Europe, Italy, and the Iberian Peninsula, which clearly originates in the MPI-ESM-LR. The reason for this skill is likely due to low frequency processes associated with large-scale phenomena, such as NAO and AMO, captured by the global model.

The reliability is quite satisfactory for both models and the predictability could be preserved but not improved by the downscaling due to reasons already mentioned above.

The predictive skill for European temperatures in winter is, in general, smaller than in summer. However, the correlation at lead years 6–10 is better than at lead years 1–5, which is counterintuitive. As speculated, slightly different representations of low frequency variability between the global model and observations could be a possible explanation.

A clear added value could be achieved by downscaling half-year precipitation sums with the regional model CCLM. Especially in summer, the regionalisation yields an increase in the predictive skill, while simultaneously improving the reliability. This is true for lead years 1–5 and lead years 6–10. The reason for the added value is the better representation of small-scale precipitation features by the regional model. However, the precipitation predictive skill over Europe shows a complex pattern, where positive and negative correlations have been found. The negative correlations originate in the global model and have been transferred to the regional scale. We speculated on several possible explanations, which could be the basis for further analysis, such as the beat frequencies and a possible problem of the zonal representation in MPI-ESM-LR.

The advantage of regionalisation for winter precipitation is slightly smaller. The positive predictive skill seems to be constrained to southern Europe, where an increase in skill is observed at the Iberian Peninsula for lead years 1–5. The predictive skill for lead years 6–10 is very weak for both models. The reliability of the hindcasts is improved by the downscaling with CCLM for all lead years and especially advantageous in mountainous regions.

Thus, dynamical downscaling is a valuable approach to improve decadal predictions of precipitation on finer spatial scales. The reason for the improvement of the predictability of precipitation is most likely that summer precipitation is a small-scale process and the dynamics of precipitation features, such as convection, are much better represented by the regional model as compared to the global model.

We have presented a first analysis of the possibilities and limitations of regional decadal predictions for Europe and shown that added value in terms of the metrics used could be achieved by regionalisation. However, a range of open questions remain: e.g. the negative correlations with observations in Central Europe; which metrics to use, e.g. terciles, contingency tables; the size of the region under study; which statistics to consider: practitioners might be more interested in the prediction of extremes like heavy precipitation, droughts, or heat waves. We will consider the prediction of extremes in a further study and expect added value through regionalisation. We also plan to study the effect of finer resolution (below 10 km). Further development stages within the MiKlip project include the production and finalisation of an improved regional ensemble system based on a new ocean initialisation of the global model (e.g. Matei et al., 2012). Fur-

ther opportunities for enhanced skill arise from combining the CCLM and REMO results to an extended multi-RCM ensemble. Additionally, an ensemble with resolution 0.44° has been produced employing annual starting dates to analyse lead time dependencies.

Acknowledgements. We acknowledge the E-OBS data set from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA&D project (<http://www.ecad.eu>).

The research programme MiKlip is funded by the German ministry of education and research (BMBF).

We also thank Andreas Weigel for fruitful discussion.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

Edited by: W. Hazeleger

References

- Bellucci, A., Gualdi, S., Masina, S., Storto, A., Scoccimarro, E., Cagnazzo, C., Fogli, P., Manzini, E., and Navarra, A.: Decadal climate predictions with a coupled OAGCM initialized with oceanic reanalyses, *Clim. Dynam.*, 40, 1483–1497, 2013.
- Berg, P., Wagner, S., Kunstmann, H., and Schädler, G.: High resolution regional climate model simulations for Germany: part I – validation, *Clim. Dynam.*, 40, 401–414, 2013.
- Chikamoto, Y., Timmermann, A., Stevenson, S., and DiNezio, P.: Decadal predictability of soil water, vegetation, and wildfire frequency over North America, *Clim. Dynam.*, under revision, 2014.
- Christensen, J. H., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., Kolli, R. K., Kwon, W.-T., Laprise, R., Rueda, V. M., Mearns, L., Menéndez, C., Räisänen, J., Rinke, A., Sarr, A., and Whetton, P.: Regional Climate Projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, 2007.
- Corti, S., Weisheimer, A., Palmer, T. N., Doblas-Reyes, F. J., and Magnusson, L.: Reliability of decadal predictions, *Geophys. Res. Lett.*, 39, 21, doi:10.1029/2012GL053354, 2012.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. Roy. Meteorol. Soc.*, 137, 553–597, 2011.

- Doblas-Reyes, F. J., Andreu-Burillo, I., Chikamoto, Y., García-Serrano, J., Guemas, V., Kimoto, M., Mochizuki, T., Rodrigues, L. R. L., and van Oldenborgh, G. J.: Initialized near-term regional climate change prediction, *Nat. Commun.*, 4, 1715, doi:10.1038/ncomms2704, 2013.
- Doms, G. and Schättler, U.: A Description of the non-hydrostatic regional model LM, Part I: Dynamics and Numerics, Tech. rep., Deutscher Wetterdienst, P.O. Box 100465, 63004 Offenbach, Germany, LM_F90 2.18, 2002.
- Dool, H. V. D.: Empirical methods in short-term climate prediction, Oxford Univ. Pr., 2007.
- Edelson, R. A. and Krolik, J. H.: The discrete correlation function: A new method for analyzing unevenly sampled variability data, *Astrophys. J.*, 333, 646–659, 1988.
- Feldmann, H., Schädler, G., Panitz, H.-J., and Kottmeier, C.: Near future changes of extreme precipitation over complex terrain in Central Europe derived from high resolution RCM ensemble simulations, *Int. J. Climatol.*, 33, 1964–1977, 2013.
- Gangstø, R., Weigel, A. P., Liniger, M. A., and Appenzeller, C.: Methodological aspects of the validation of decadal predictions, *Clim. Res.*, 55, 181–200, doi:10.3354/cr01135, 2013.
- García-Serrano, J., Doblas-Reyes, F. J., Haarsma, R. J., and Polo, I.: Decadal prediction of the dominant West African monsoon rainfall modes, *J. Geophys. Res.-Atmos.*, 118, 5260–5279, 2013.
- Giorgi, F., Jones, C., and Asrar, G.: Addressing climate information needs at the regional level: the CORDEX framework, *World Meteorol. Organ (WMO) Bull.*, 58, 175–183, 2009.
- Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., Kharin, V., Merryfield, W., Deser, C., Mason, S., Kirtman, B., Msadek, R., Sutton, R., Hawkins, E., Fricker, T., Hegerl, G., Ferro, C., Stephenson, D., Meehl, G., Stockdale, T., Burgman, R., Greene, A., Kushnir, Y., Newman, M., Carton, J., Fukumori, I., and Delworth, T.: A verification framework for interannual-to-decadal predictions experiments, *Clim. Dynam.*, 40, 245–272, 2013.
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.-Atmos.*, 113, D20119, doi:10.1029/2008JD010201, 2008.
- Hurrell, J., Visbeck, M., and Pirani, A.: WCRP Coupled Model Intercomparison Project – Phase 5, Special Issue of the CLIVAR Exchanges Newsletter, available at: http://www.clivar.org/publications/exchanges/Exchanges_56.pdf, (last access: 1 September 2013), 2011.
- Jacob, D.: A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin, *Meteorol. Atmos. Phys.*, 77, 61–73, 2001.
- Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: new high-resolution climate change projections for European impact research, *Regional Environmental Change*, 1–16, 2013.
- Keenlyside, N. S., Latif, M., Jungclaus, J., Kornblueh, L., and Roeckner, E.: Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, 453, 84–88, 2008.
- Latif, M., Delworth, T., Dommenges, D., Drange, H., et al.: Dynamics of decadal climate variability and implications for its prediction, in: *Proc OceanObs09: sustained ocean observations and information for society*, edited by: Hall, J., Harrison, D. E., and Stammer, D., no. 2 in Publication WPP-306, ESA, 2010.
- Mason, S. J.: From dynamical model predictions to seasonal climate forecasts. *Seasonal climate variability: Forecasting and managing risk*, Springer Netherlands, 2008.
- Matei, D., Pohlmann, H., Jungclaus, J., Müller, W., Haak, H., and Marotzke, J.: Two tales of initializing decadal climate prediction experiments with the echam5/mpicom model, *J. Climate*, 25, 8502–8523, 2012.
- Mieruch, S., Schröder, M., Noël, S., and Schulz, J.: Comparison of decadal global water vapor changes derived from independent satellite time series, *J. Geophys. Res.-Atmos.*, doi:10.1002/2014JD021588, 2014.
- Müller, W. A., Baehr, J., Haak, H., Jungclaus, J. H., Kröger, J., Matei, D., Notz, D., Pohlmann, H., von Storch, J. S., and Marotzke, J.: Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology, *Geophys. Res. Lett.*, 39, L22707, doi:10.1029/2012GL053326, 2012.
- Palmer, T. N., Doblas-Reyes, F. J., Weisheimer, A., and Rodwell, M. J.: Toward seamless prediction: Calibration of climate change projections using seasonal forecasts, *Bull. Amer. Meteor. Soc.*, 89, 459–470, 2008.
- Panitz, H.-J., Dosio, A., Büchner, M., Lüthi, D., and Keuler, K.: COSMO-CLM (CCLM) climate simulations over CORDEX-Africa domain: analysis of the ERA-Interim driven simulations at 0.44° and 0.22° resolution, *Clim. Dynam.*, 42, 11–12, 3015–30338, doi:10.1007/s00382-013-1834-5, 2014.
- Pohlmann, H., Jungclaus, J. H., Köhla, A., Stammer, D., and Marotzke, J.: Initializing decadal climate predictions with the gecco oceanic synthesis: effects on the north atlantic, *J. Climate*, 22, 3926–3938, 2009.
- Räsänen, J. and Ylhäisi, J. S.: How much should climate model output be smoothed in space?, *J. Climate*, 24, 867–880, 2011.
- Rockel, B., Will, A., and Hense, A.: The regional climate model COSMO-CLM (CCLM), *Meteorol. Z.*, 17, 347–348, 2008.
- Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American winters, *Geophys. Res. Lett.*, 41, 2514–2519, 2014a.
- Scaife, A. A., Athanassiadou, M., Andrews, M., Arribas, A., Baldwin, M., Dunstone, N., Knight, J., MacLachlan, C., Manzini, E., Müller, W. A., Pohlmann, H., Smith, D., Stockdale, T., and Williams, A.: Predictability of the quasi-biennial oscillation and its northern winter teleconnection on seasonal to decadal timescales, *Geophys. Res. Lett.*, 41, 1752–1758, 2014b.
- Smith, D., Scaife, A., Boer, G., Caian, M., Doblas-Reyes, F., Guemas, V., Hawkins, E., Hazeleger, W., Hermanson, L., Ho, C., Ishii, M., Kharin, V., Kimoto, M., Kirtman, B., Lean, J., Matei, D., Merryfield, W., Müller, W., Pohlmann, H., Rosati, A.,

- Wouters, B., and Wyser, K.: Real-time multi-model decadal climate predictions, *Clim. Dynam.*, 41, 2875–2888, 2012.
- Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, *Adv. Model. Earth Syst.*, 5, 146–172, doi:10.1002/jame.20015, 2013.
- Thiebaut, H. J. and Zwiers, F. W.: The interpretation and estimation of effective sample size, *J. Climate Appl. Meteor.*, 23, 800–811, 1984.
- Trail, M., Tsimpidi, A. P., Liu, P., Tsigaridis, K., Hu, Y., Nenes, A., and Russell, A. G.: Downscaling a global climate model to simulate climate change over the US and the implication on regional and urban air quality, *Geosci. Model Dev.*, 6, 1429–1445, doi:10.5194/gmd-6-1429-2013, 2013.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, I., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, *Q. J. Roy. Meteorol. Soc.*, 131, 2961–3012, 2005.
- van Oldenborgh, G. J., Doblas-Reyes, F. J., Wouters, B., and Hazeleger, W.: Skill in the trend and internal variability in a multi-model decadal prediction ensemble, *Clim. Dynam.*, 38, 1263–1280, 2012.
- von Storch, H. and Zwiers, F. W.: Testing ensembles of climate change scenarios for “statistical significance”, *Climatic Change*, 117, 1–2, 1–9, doi:10.1007/s10584-012-0551-0, 2013.
- Wagner, S., Berg, P., Schädler, G., and Kunstmann, H.: High resolution regional climate model simulations for Germany: Part II—projected climate changes, *Clim. Dynam.*, 40, 415–427, 2013.
- Weigel, A. P., Liniger, M. A., and Appenzeller, C.: Seasonal ensemble forecasts: Are recalibrated single models better than multi-models?, *Mon. Weather Rev.*, 137, 1460–1479, 2009.
- Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences*, Academic Press, 3rd Revised Edn., 2011.
- Wu, Z., Huang, N. E., Long, S. R., and Peng, C.-K.: On the trend, detrending, and variability of nonlinear and nonstationary time series, *Proc. Natl. Aca. Sci.*, 104, 14889–14894, 2007.