Geosci. Model Dev. Discuss., 6, 5711–5745, 2013
www.geosci-model-dev-discuss.net/6/5711/2013/
doi:10.5194/gmdd-6-5711-2013

Geoscientific Model Development Discussions
Open Access

The regional MiKlip decadal forecast ensemble for Europe

S. Mieruch[1], H. Feldmann[1], G. Schädler[1], C.-J. Lenz[2], S. Kothe[3], and C. Kottmeier[1]

[1]Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany
[2]Deutscher Wetterdienst, Offenbach, Germany
[3]Institute for Atmospheric and Environmental Sciences, Goethe-University Frankfurt, Frankfurt am Main, Germany

Received: 30 September 2013 – Accepted: 4 November 2013 – Published: 22 November 2013

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Funded by the German Ministry for Education and Research (BMBF) a major research project called MiKlip (Mittelfristige Klimaprognose, Decadal Climate Prediction) was launched and global as well as regional predictive ensemble hindcasts have been generated. The aim of the project is to demonstrate for past climate change whether predictive models have the capability of predicting climate on time scales of decades. This includes the development of a decadal forecast system, on the one hand to support decision making for economy, politics and society for decadal time spans. On the other hand, the scientific aspect is to explore the feasibility and prospects of global and regional forecasts on decadal time scales. The focus of this paper lies on the description of the regional hindcast ensemble for Europe generated by COSMO-CLM and on the assessment of the decadal variability and predictability against observations. To measure decadal variability we remove the long term bias as well as the long term linear trend from the data. Further, we applied low pass filters to the original data to separate the decadal climate signal from high frequency noise. The decadal variability and predictability assessment is applied to temperature and precipitation data for the summer and winter half-year averages/sums. The best results have been found for the prediction of decadal temperature anomalies, i.e. we have detected a distinct predictive skill and reasonable reliability. Hence it is possible to predict regional temperature variability on decadal timescales, However, the situation is less satisfactory for precipitation. Here we have found regions showing good predictability, but also regions without any predictive skill.

1 Introduction

Interest in longer term climate predictions in the range of about 10 yr is growing. Such predictions, as opposed to projections, would be very useful for all branches of public life/activity and planning purposes, e.g. in agriculture, energy management, hydrology, and health. In the sense of seamless prediction (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 and it is not known to what extent useful predictions are possible in terms of lead time, geographical position, spatial resolution, meteorological variable and statistics like means or extremes. There is, however, widespread agreement about some necessary requirements for such predictions to be successful: (i) decadal predictions need to be produced by coupled atmosphere-ocean models, (ii) predictability lies mainly in the slow components of the climate system, i.e. the oceans and the soil. Skillful modeling and good initialization of these components is essential (Keenlyside et al., 2008), (iii) predictability must come from the large scale processes and interactions (like AMO, ENSO, quasi-biennial oscillation), who must be captured by the global models, and (iv) assuming that the models capture the effects of external forcing (especially concentration changes of greenhouse gases), prediction means essentially prediction of long term (decadal) internal variability. This is a mixed dynamical, in the sense of deterministic chaos (Lorenz, 1963), and stochastic quantity, so therefore ensembles of simulations are required. The effects of initialization have been discussed e.g. by Keenlyside et al. (2008); Pohlmann et al. (2009); Müller et al. (2012); Smith et al. (2012).

For practical 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 provide in many cases added value (Feldmann et al., 2013; Wagner et al., 2013; Berg et al., 2013; Trail et al., 2013); if such added value can also be found in regionally downscaled predictions is presently an open question. It is also an open question which metrics to use to measure the skill of the predictions and which metrics are useful for applications. Whereas science is interested in variability and ensemble metrics (Goddard et al., 2013; Gangstø et al., 2013), practitioners require above/below climatology, return values, frequency and duration of extremes (which is difficult due to bias) with high spatial resolution. The German Ministry for Education and Research (BMBF) has launched a major program 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 (Stevens et al., 2013) and the regional climate models COSMO-CLM (Rockel et al., 2008; Panitz et al., 2013) and REMO (Jacob, 2001) for regional downscaling. The project consists of five modules dealing with the different aspects of predictability described above: initialization, 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.

The focus of this paper is on the description and the assessment of regional predictions with COSMO-CLM. Section 2 briefly describes the setup of the MPI-ESM simulations and gives an overview over the setup of the COSMO-CLM simulations, the construction of the ensemble and the data used for validation. Section 3 describes the 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 of the results is given in Sect. 5, conclusions and an outlook are given in Sect. 6.

2 Experimental design – construction of the regional decadal ensemble

The intent of MiKlip is to develop and improve a decadal prediction system in several development stages. In a first phase a so called "baseline ensemble" of decadal predictions has been established. It encompasses the global decadal simulations performed with the MPI-ESM (Stevens et al., 2013) for CMIP5 (Hurrell et al., 2011) with an ensemble size of 10 members. This baseline ensemble is used as a starting point and basis for further developments within MiKlip. With respect to the potential users of such information, regional downscaling of these global predictions is – after validation of the GCM results – an important next step. Among the regions selected for this regional downscaling is Europe. Two regional climate models (RCMs) – namely COSMO-CLM (Consortium of small scale modeling in climate mode; CCLM hereafter) and REMO have been applied. 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. The focus for this paper is on the simulations with CCLM. This RCM is used in the same model version and as for CORDEX (Panitz et al., 2013).

The global baseline ensemble with the MPI-ESM in low resolution (MPI-ESM-LR) consists of 10 realizations with starting years every 5 yr between 1960 until 2000 and annual initialization after 2000 according to the CMIP5 protocol (Taylor et al., 2012, decadal1960: 1 January 1961–31 December 1970; decadal1965: 1 January 1966–31 December 1975; ...) and three ensemble members for the in-between-years. The ensemble uses 1 day time-lagged initialization of the atmosphere and an anomaly initialization for the ocean (Müller et al., 2012; Matei et al., 2012). The atmospheric resolution is T63 ($\approx 1.86°$) horizontally with 47 vertical levels up to 0.1 hPa in the vertical. The resolution of the ocean component is 1.5° on average.

For the first regional ensemble a larger ensemble size was preferred to a higher number of starting dates. This is a reasonable choice to analyze the spread and reliability of the regional ensemble with respect to the global ensemble. On the other hand a model climatology for the whole period 1961–2010 is necessary to calculate the model anomalies. Therefore, all 10 available realizations of the MPI-ESM-LR for 5 starting dates (decadal1960–decadal2000) were downscaled covering the whole 50 yr period.

A CCLM reference simulation is performed with re-analysis forcing. This simulation starts in 1959 using ERA40 (Uppala et al., 2005) as initial and boundary conditions. The first two years are used as spin-up time for the slowly varying lower boundary of the model. From 1979 until 2010 ERA Interim-forcing (Dee et al., 2011) is applied. The soil initial conditions for the GCM-driven CCLM simulations were derived from

GMDD
6, 5711–5745, 2013

this reference simulation. This has the advantage that the soil variables are as consistent as possible with observations, provided that the soil inventories and SVATs (Soil–Vegetation–Atmosphere Transfer schemes) used in GCM and RCM are consistent. This can be regarded as an anomaly initialization of the RCM, avoiding drifts and inconsistent surface fluxes in the initial phase of the predictions (Khodayar et al., 2013).

The E-OBS v8.0 climatology (Haylock et al., 2008) for near-surface temperature and precipitation was used to evaluate the model performance. E-OBS is a gridded observational dataset 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/) information.

3 Data pre-processing and skill metrics

An important aspect of assessing predictability, such as skill (Goddard et al., 2013) and reliability (Corti et al., 2012) of decadal hindcasts vs. observations is to decide which data pre-processing and metrics to use to characterize decadal variability. Different approaches of pre-processing are followed; Bellucci et al. (2013) for instance analyse 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 confront initialized forecasts with uninitialized projections. Similar to Latif et al. (2010) (and in parts Bellucci et al., 2013) we decided to remove the long term means and trends from the time series to analyse decadal variability to avoid the difficulties interpreting a mixture of long term and decadal changes. Furthermore, also the long term trend (50 yr) is just a snapshot on much larger time scales such as centuries and millennia as e.g. modelled by Friedrich et al. (2010). To address the problem of decadal variability, a single E-OBS anomaly time series of summer half-year temperature means near Łódź (Poland) is shown exemplarily in Fig. 1. The detrended/unfiltered data are shown as a thin line. The long term mean and trend

GMDD
6, 5711–5745, 2013

have been removed, thus the anomalies are centered around zero. We see high summer to summer variability and these high frequency fluctuations cannot be predicted using decadal model initializations. To extract the potentially predictable signal we applied a 9 yr moving average low pass filter to the data (just for visualization), which is plotted as a thick line, where we have shaded the area enclosed with the curve and the zero line in Fig. 1. This line represents what we understand by decadal variability, it is the part of the climate varying on decadal time scales. Now we have separated the decadal variability from the fast interannual fluctuations and filtering makes it easier to identify anomalies on decadal time scales. The 60's were an anomalous warm decade, the 70's were colder, whereas the 80's and 90's were more medium decades and finally the 00's were a warmer than normal decade. It has to be kept in mind that the long term trend has been removed, which amounts for this example (Fig. 1) to about $0.4\,^{\circ}\mathrm{C\,decade}^{-1}$, hence absolutely the 90's were warmer than the 60's, but not with respect to the detrended temperature. After removing the long term means and trends from the CCLM hindcasts and the E-OBS observations a second pre-processing step is applied to both, the smoothing of the data with 5 and 9 yr moving average filters. It is important to note that the filtering has been performed only within decades, i.e. without shifting the moving filter window over decadal boundaries since no observations should be used from a future decade to increase the predictability of the present decade. The following table (Table 1) shows on which years the averaged points have been based on.

Smoothing of the data is beneficial in skill assessment due to reduction of the unpredictable grid-scale noise (Räisänen and Ylhäisi, 2011). Goddard et al. (2013) advocate a 5° latitude × 5° longitude spatial smoothing for precipitation and a 10° latitude × 10° longitude smoothing for temperature. Further they provide analyses on different time scales, i.e. year 1, years 2–5, years 6–9 and years 2–9, which is "... somewhat arbitrary, but it represents a small set of cases that can illustrate the quality of the information for different lead times and temporal averaging." In contrast to Goddard et al. (2013), who deal with global data, we focus on regionalized data and we analyse the

GMDD
6, 5711–5745, 2013

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

data at their original 25 km resolution without spatial smoothing. Similar to Goddard et al. (2013) we use temporal smoothing, however as explained above, we use moving average filters instead of averaging over selected time ranges. The reason why we prefer the moving average filter is that the number of data points is not decreased and it is accounted for changes within an averaging range e.g. a decade. However, moving average filtering introduces strong autocorrelations, which have to be considered when performing significance analyses.

After the pre-processing the following metrics will be used to characterize the CCLM ensemble vs. observations on decadal time scales:

- Skill:
  To quantify the predictive skill of the regional CCLM ensemble against observations we explore the Pearson correlation coefficient ρ applied to anomalies (also known as ACC (anomaly correlation coefficient); e.g. Bellucci et al., 2013). If we denote the CCLM ensemble mean at a specific location i as $m_{t,i}$ and the corresponding observations as $o_{t,i}$, where $t = 1, \ldots, N$ represents the time index with in total $N = 50$ data points (semi-annual means 1961–2010), the correlation coefficient is given by

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

  where it has to be noted again that $m_{t,i}$ and $o_{t,i}$ are anomaly time series, which have been bias and long term trend adjusted. The interpretation of the value of the correlation coefficient is very individual and depends strongly on the underlying experiment and data. A correlation coefficient of 0.8 can be very low regarding e.g. high performance experiments to verify a physical constant. Otherwise 0.8 can be very large e.g. in social sciences, where complicated factors interfere. For our decadal predictions we stick to the common interpretation that a correlation coefficient up to 0.3 is regarded as low, coefficients from 0.5 are good and a ρ larger than 0.7 indicates a very strong dependence.

GMDD
6, 5711–5745, 2013
Decadal forecast
S. Mieruch et al.

As shown by Kharin and Zwiers (2003), the correlation coefficient can be used to estimate the so called potential predictability, i.e. the ratio of the variance of the potentially predictable signal and the total observed variance. It is called potential, because it represents the level of skill, which can be attained when the long term means and trends have been removed (Murphy, 1988). Moreover, the correlation coefficient is invariant to a change in the location and scale of a variable X, i.e. a transform in the sense of $a+bX$. Thus, ρ measures rather the tendencies of the anomalies and not the actual agreement.

Further, we have estimated the statistical significance of the correlation coefficient on grid point basis with the test statistic

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

on a t-distribution with N_{eff} degrees of freedom. Due to the low pass filtering strong autocorrelations are introduced to the time series, thus the effective number of degrees of freedom is reduced. Therefore we account for autocorrelations according to (Wilks, 2006)

$$N_{\mathrm{eff}} = N \cdot \frac{1-\phi}{1+\phi}, \tag{3}$$

where ϕ is the autocorrelation at lag 1 of $[m_{t,i} \cdot o_{t,i}]$. Although only lag 1 is considered explicitely, further autocorrelations at higher lags are also considered implicitly, since the correction approach is based on autoregressive processes of order 1 (AR[1]) (Thiebaux and Zwiers, 1984), whose autocorrelation function decreases exponentially, thus considering also higher lags than 1.

– Fidelity:
DelSole and Shukla (2010) coined the term “fidelity” and defined it as a measure

GMDD
6, 5711–5745, 2013

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◂ | ▸
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

of the agreement between model and observational climatological distributions. They exploited metrics from information theory (Shannon, 1948) such as the relative entropy to measure fidelity, which is quite appealing. However, because of the simplicity and the greater acceptance we use the well known χ^2 test to compare model and observational distributions. The χ^2 test requires the choice of the number of bins used to estimate the distribution, which is somewhat subjective. For our study we chose the number of bins to be 6, because it is a good compromise between the number of data points per time series (50) and the resolution of the distribution. The χ^2 test statistic is estimated on grid pixel basis and a p value is derived, which gives the probability of accepting the null hypothesis of equal model and observational distributions. Since the p values are derived on grid pixel basis for all 10 ensemble members, a total of 10 p values is gathered at each grid point and then averaged for presentation.

– 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 e.g. Mason, 2008). As mentioned by Weigel et al. (2009) a normally distributed ensemble is reliable if and only if the rmse of the ensemble mean and the observations is identical to the time-mean ensemble spread $\sqrt{\left\langle \sigma_{\mathrm{ens}}^2 \right\rangle}$. The ensemble is called underconfident if $\mathrm{rmse}(\mu, x) < \sqrt{\left\langle \sigma_{\mathrm{ens}}^2 \right\rangle}$, it is called overconfident if $\mathrm{rmse}(\mu, x) > \sqrt{\left\langle \sigma_{\mathrm{ens}}^2 \right\rangle}$, and calibrated if $\mathrm{rmse}(\mu, x) = \sqrt{\left\langle \sigma_{\mathrm{ens}}^2 \right\rangle}$. Loosely speaking, reliability measures if the ensemble spread covers the model errors. Hence,

GMDD
6, 5711–5745, 2013
Decadal forecast
S. Mieruch et al.

Weigel et al. (2009) define reliability as

$$\mathrm{REL}_i = \frac{\mathrm{rmse}(\mu_{t,i}, x_{t,i}) - \sqrt{\left\langle \sigma^2_{i,\mathrm{ens}} \right\rangle_t}}{\mathrm{rmse}(\mu_{t,i}, x_{t,i})}, \tag{4}$$

where we use the index i to indicate a single grid point and t to depict the time index.

4 Decadal variability assessment

In Sect. 3 we have defined what we understand by decadal variability in this study (cf. Fig. 1). To quantify the predictability of decadal variability we explore three characteristics of the ensemble, which are the skill, the fidelity and the reliability. We measure these characteristics using the correlation coefficient ρ (skill), the χ^2 test (fidelity) and a metric based on the rmse and ensemble spread (reliability) (cf. Sect. 3). We have chosen these three measures because these are the simplest metrics to characterize the most important aspects of a forecast ensemble. Another advantage of these simple metrics is that results shown here are easily to understand and easily reproducible. The findings for the predictability of summer temperature are the easiest ones to interpret, thus we decided to discuss these results in detail and show the corresponding figures below. The results for winter temperature and summer and winter precipitation are only shortly discussed and the corresponding figures can be retrieved in the Supplement.

Before we analyse the CCLM ensemble for the whole of Europe we will exemplarily show summer temperature results on three different locations in Europe in Fig. 2. The top left panel of Fig. 2 depicts the data at a grid point near Łódź in Poland. The yearly unfiltered E-OBS (lilac) and CCLM ensemble mean (orange) anomalies are shown as thin lines, whereas the thick lines represent 5 yr moving average filters. Additionally, the CCLM ensemble spread, i.e. the standard deviation over the ensemble for each time step, is shown as a grey shaded area. Due to the decadal initialization of CCLM in 1961, 1971, 1981, 1991 and 2001 we have separated the decades from each other. The correlation coefficient for the original yearly data in the top left panel in Fig. 2 is 0.33. This correlation can be increased by applying the 5 yr moving average filter to 0.69 and even further increased to 0.84 using a 9 yr moving average filter, as can be seen in the top right panel of Fig. 2. Looking again at the results shown in the top right panel of Fig. 2 we observe a strong predictive skill and it is clear from visual inspection that the CCLM ensemble mean is able to capture the tendency of the decadal variability. We will now look at different periods within the decades, starting with decadal beginnings, i.e. 1961, 1971, 1981, 1991, 2001, which are the dates where the retrospective forecasts have been carried out. The forecast from 1961 predicts an anomalous warm decade with a decadal trend to the "normal" state at the end of the decade, which is in agreement with the E-OBS data. Moving further on in time, the model has been initialized anew in 1971 and CCLM predicts a further continuing decreasing decadal trend and the tendency that the 70's are an anomalous cold decade are well captured. The next two decades can be characterized as normal without strong decadal variations (with respect to the long term trend) and have been well forecasted by CCLM. Finally the 00 decade has been correctly forecasted as an anomalous warm period. If we look at the thin curves in Fig. 2 we see that it is not possible to forecast single years, but applying a low pass filter (thick curves) we can successfully predict the tendency on decadal time scales. Most, but not all regions in Europe show such good decadal predictability for summer temperature as observed in Poland. A slightly worse decadal predictability has been found e.g. in Karlsruhe Germany, shown in the middle panels of Fig. 2. Here, the correlation coefficient is increased with low pass filtering to 0.53 (5 yr, left panel) and 0.72 (9 yr, right panel). Unsatisfactory results are found near Rome, Italy (Fig. 2 bottom panels). The correlation coefficient is negligible small at about 0.1, thus decadal predictive skill is absent. To interpret the observed correlations, we will calculate, as an example, the hit rate of a binary prediction of a decade being above or below the average (trend) in the sense of a categorical forecast. The hit rate (shown in Table 2) is

GMDD
6, 5711–5745, 2013

given by the fraction of correctly forecasted decades (above/below zero) and the total number of decades in percent. It has to be noted that for the here used anomalies a hit rate of 50 % is expected simply by guessing. A detailed view on the whole of Europe will be discussed in the following.

4.1 Summer temperature

4.1.1 Bias and trend

The long term means from 1961 to 2010 of average summer temperatures are shown for E-OBS in Fig. 3 in the top left panel. The temperature averages for the CCLM ensemble mean are shown in the top middle panel and the difference CCLM minus E-OBS is shown in the top right panel of Fig. 3. The three bottom panels of Fig. 3 show the long term trends (1961–2010) of E-OBS and CCLM (ensemble mean) summer temperatures as well as the trend differences (bottom right panel).

With respect to the pre-processing, our analysis of the decadal predictability is independent of the long term means and trends. However, for completeness we will briefly discuss the long term performance of the CCLM decadal hindcasts without judging too much on the quality of reproducing long term means and trends, which would be the subject to another analysis. Regarding the summer temperature means (top row of Fig. 3) of E-OBS and CCLM we see that the patterns agree well. A known cold bias of −2 to 0 °C exists (Berg et al., 2013), which is quite homogeneously distributed yet slightly larger in the alpine region. The grainy structure seen in mountainous regions, especially the Alps, the Pyrenees, and the northern coast of Norway, can be attributed to orographic differences between E-OBS and CCLM. Thus, for a thorough analysis of the bias a simple orography correction applied to CCLM would remove the grainy structures.

The CCLM long term trends (bottom middle panel of Fig. 3) are very weak, however the general gradient from smaller trends in northern Europe to larger trends in southern Europe is captured. The bottom right panel of Fig. 3 shows a few very localized spots of

GMDD
6, 5711–5745, 2013

large differences, e.g. in the region Auvergne (France), Romania and Moldavia, which are most probably artefacts in the E-OBS data.

4.1.2 Skill

Figure 4 shows the results of the correlation coefficient for Europe. The left panel of Fig. 4 presents the correlation for the original unfiltered anomalies of temperature summer half-year means. Most of the regions show a positive correlation with a pronounced maximum at the Iberian Peninsula and eastern Europe of up to 0.4. Significant correlation (cf. Eq. 2) is indicated by stippling, which means that the probability (p value) of obtaining such large correlations, assuming that the null hypothesis of zero correlation is true, is < 0.05.

However, also regions of very low or even negative correlations are found, e.g. in Italy, Southwest France, Sweden, Norway and the Balkan region. To extract the predictable signal from the high frequency year to year fluctuations of the anomalies we have applied centered moving average filters of window size 5 and 9 yr to the CCLM ensemble mean and the E-OBS observations as described in Sect. 3. The application of the moving average low pass filter increases the correlation coefficient, which is shown in Fig. 4, but simultaneously smears out the temporal resolution. That means that statements on predictability can only be made on the basis of larger temporal periods. Nevertheless, the weak correlation of the unfiltered time series indicates clearly that the claim on predicting single temperature anomaly summers is too ambitious. Thus averaging seems to make sense. The middle panel of Fig. 4 shows the correlation coefficient for the 5 yr moving average filter and large parts of the former negative correlated regions become positively correlated (France, Sweden) and additionally significant (stippling). However, the correlation coefficient at the Iberian peninsula is only slightly increased and the significance is lost due to the strong autocorrelations introduced by filtering. That means, filtering is not always advantageous. Further, the results in parts of Italy, a small area in Sweden and the Balkan have not been improved. Increasing the moving average filter window size to 9 yr (right panel of Fig. 4) increases also the correlation coefficient and the significance in most European regions. However, there is no clear improvement at the Iberian Peninsula, Italy and the Balkan. Thus, filtering can crucially improve the predictability like e.g. in Poland and eastern Europe, but it can also be effectless or even decrease the predictability. This can be understood for instance if there is such a weak potentially predictable signal that it cannot be extracted from the high frequency noise. In such a case filtering is effectless or could even yield a worse situation.

4.1.3 Fidelity

As described in Sect. 3 the fidelity is measured using the χ^2 test and a p value, which gives the probability of accepting the null hypothesis of equal model and observational distributions, is plotted in Fig. 5. As can be seen from Fig. 5 the averaged p values are mostly large, which is an indication to accept the null hypothesis of equal model and observational distributions. The application of moving average filter improves the situation and the p values are getting larger. Thus, it can be concluded that the summer CCLM and E-OBS climatological distributions are not significantly different.

4.1.4 Reliability

Figure 6 shows the reliability results for Europe for the original unfiltered data in the left panel. Over most regions, the reliability is very satisfactory and the ensemble spread covers about 80 % to 120 % of the rmse, i.e. $-0.2 < \mathrm{REL} < 0.2$ (a calibrated ensemble would yield $\mathrm{REL} = 0$). It is interesting to see that the British Isles, Scandinavia and the northern part of Europe, along the coast yield slightly overconfident reliability, whereas the rest of Europe show more underconfident results. This observation is most probably connected to the maritime influence in North Europe. Slightly more underconfident results have been observed in eastern Europe. The application of moving average filter do not improve the situation, and the reliability covers the range of $-0.4 < \mathrm{REL} < 0.4$ for the 9 yr filter. Further a change in the observed reliability pattern has been observed due to the filtering. It is interesting to see that for the 5 and 9 yr filters northern and middle Europe become underconfident, whereas southern Europe becomes overconfident. Remembering the correlation results (Fig. 4), which have been improved by filtering, we observed the contrary here, i.e. a worsening of the reliability utilising the moving average filters. Thus, with respect to the effect of the moving average filter, the correlation and reliability seem to act in opposite directions. Finally if one thinks about a real future forecast, a decision in the form of a compromise cannot be avoided, i.e. the compromise between good correlation and good reliability.

4.2 Winter temperature

Similar to the good results found for summer temperature, decadal variability can be forecasted well for winter temperatures for a large part of Europe. The Fig. S1 shows the results of the analysis of the decadal CCLM winter temperature ensemble vs. E-OBS observations. The bias looks similar to the summer temperature bias (Fig. 3) and ranges from −2 to 0 °C, which is known (Berg et al., 2013). The observed long term trend patterns of CCLM and E-OBS agree well with large positive trends in northern Europe and weak positive trends in southern Europe. However, the CCLM long term trends are too weak and a difference in the order of −2 to $0\,\mathrm{K}\,(50\,yr)^{-1}$ has been found.

The correlation coefficient of the decadal variability can be improved significantly by filtering out the year to year fluctuations. Without filtering the correlation coefficient is very small, about −0.2 to 0.2. The moving average filter of window size 5 yr reveals an increase in the correlation coefficient for a large part of Europe to up to 0.6. Applying the 9 yr moving average filters further increases the correlation coefficient especially in south east Europe. These correlations are significant in a statistical sense (stippling). However, the correlation coefficient has not been improved at the Iberian Peninsula, Scandinavia, the Baltic states, Belarus and Ukraine. The pattern of the increase of ρ due to the filtering covers Europe like a large swath from northwest to southeast. This could be a hint to the causes of the decadal predictability observed here, arising from a north-atlantic large scale low frequency variability.

The fidelity of CCLM winter temperature anomaly ensemble members and E-OBS indicates no significant differences between the distributions. Filtering improves the situation and very high p values have been observed. Thus the null hypothesis of equal distributions between model and observations has to be accepted.

The reliability shows quite reasonable results and lies inbetween −0.2 to 0.2. The northern part of Europe is slightly overconfident, i.e. the ensemble spread is too small, whereas the southern part shows slightly underconfident covering of the uncertainty. It seems that the reliability of winter temperatures is not so susceptible to the filtering and changes only very slightly.

4.3 Summer precipitation

Forecasting precipitation is more challenging than forecasting temperature, because of the stochastic dynamics of rain. Precipitation events can locally be very different, extreme, intermittent and stochastic. Therefore regarding decadal forecasts, it is expected that the influence of large scale processes is weaker on precipitation than on temperature. Obviously, internal dynamics dominate short term precipitation, but as for the decadal analysis of temperature, we will assess the predictive skill and reliability of precipitation on summer and winter sums and filtered time series to reduce the influence of short term fluctuations. Figure S2 in the Supplement shows the results of the decadal variability assessment on summer precipitation. The wet bias mostly in the order of 60 mm (more than 200 mm in the alpine region) is large, but a known issue of CCLM (Berg et al., 2013). However, the wet bias is unproblematic for our approach, because it is removed for subsequent analyses. The precipitation trends are patchy for E-OBS and the significance of these trends is questionable. However, the trend patterns of CCLM and E-OBS largely agree, showing positive trends in northern Europe and negative trends in the southern parts.

The correlation shows a patchy structure, which is expected for precipitation. For the unfiltered time series, positive, yet non-significant correlations are found in Scandinavia and eastern Europe. However, about 7 significant spots can be seen in eastern Europe

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

GMDD

6, 5711–5745, 2013

Decadal forecast

S. Mieruch et al.

and one significant spot in Scandinavia. The causes of good correlation at these spots are not clear. Significant negative correlations are found in south east France. Filtering improves the situation slightly in most regions. The effect is stronger e.g. in Great Britain and Norway. However the correlations remain weak and decadal forecasts for precipitation should be taken with care.

The distributions of summer precipitation of CCLM and E-OBS are not significantly different, as indicated by large p values in the fidelity plots.

Regarding the reliability, the CCLM ensemble is underconfident in most cases. However, the reliability is still reasonable around -0.2. This effect of underconfidence increases with low pass filtering.

4.4 Winter precipitation

Similar to summer precipitation, a known winter precipitation (Fig. S3) wet bias exists (Berg et al., 2013). This bias is stronger in south Europe than in the northern parts, except Norway. Consistent with the above observations, the CCLM long term trends are too weak. However the patterns agree with E-OBS with negative trends in southern Europe and positive trends in northern Europe.

The correlation coefficient is small for all regions. A large part of Europe shows small negative correlations for the unfiltered data. Only on the Iberian Peninsula and in south east Europe larger areas of positive correlations are retrieved. With low pass filtering, the positive correlations in the south eastern parts of Europe are increased, but also negative trends e.g. in northern Germany are amplified. Thus, it is questionable if low pass filtering is an advantage here.

Finally, the reliability is mostly underconfident as in the case for summer precipitation, however reasonable small (≈ -0.2), except some very localised spots.

GMDD
6, 5711–5745, 2013
Decadal forecast
S. Mieruch et al.

5 Summary

In this study we have assessed the predictive skill of decadal variability of summer and winter temperatures and precipitation in Europe. For this purpose we generated a 10 member ensemble of climate simulations (1961–2010) with the regional model CCLM at 25 km resolution, driven by decadal predictions from the global model MPI-ESM. These model runs have been initialised every 10 yr from 1961 to 2001.

A successful prediction of decadal variability could be very useful for society and all fields of human activity. It is also a scientific challenge to investigate the potential of decadal predictions, hence the initiation of the MiKlip program, which aims at the development of a decadal prediction system.

Decadal potential predictability has been found for summer temperatures in most regions in Europe. Here, low pass filtering is beneficial and increases the predictability at the expense of the temporal resolution of the forecast. On decadal time scales, for the 9 yr moving average filter, we found significant correlations between CCLM temperatures and observations above 0.6 in northern Europe and up to 0.6 in the southern part (non-significant).

The fidelity of individual summer temperatures ensemble members forecasted by the CCLM and E-OBS observations indicates no significant differences between their climatological distributions. In contrast, large p values have been retrieved by the χ^2 test, which definitely supports the acceptance of the null hypothesis of equal distributions. This effect is enhanced for most regions in Europe by low pass filtering.

The reliability of summer temperatures is reasonable ($-0.2 < \text{REL} < 0.2$) and the ensemble spread covers 80 % of the rmse between ensemble mean and observations in the case of overconfidence. In the case of underconfidence the ensemble spread is only slightly too large (20 % larger than the rmse). However, low pass filtering worses the reliability to about $-0.6 < \text{REL} < 0.6$ with positive values (overconfident) in southern Europe and negative values (underconfident) in northern Europe. Thus, a kind of conflict emerges regarding the low pass filtering between the correlation coefficient and the

GMDD
6, 5711–5745, 2013

reliability. The correlation coefficient can be increased by low pass filtering, which on the other hand degrades the reliability. Therefore any pre-processing of the data, such as filtering, depends on the application and intention of the forecast. In the sense of categorical forecasts, e.g. below/above the mean, averaging over whole decades (which is essentially low pass filtering) is the basis of such an analysis. If we were interested in the distribution of extremes, low pass filtering would be the wrong pre-processing, because it would smear out any extreme values.

The assessment of decadal variability of winter temperature reveals that a predictive skill of the regional ensemble exists here as well. Low pass filtering increases the correlation coefficient up to 0.5–0.7, which shows an interesting pattern like a large swath over Europe from the northwest to the southeast except for the Iberian Peninsula and Scandinavia. The reason for such a pattern could be e.g. a large scale atmospheric teleconnection.

The prediction of total precipitation is generally a challenge and much harder than temperature forecasts. As a starting point we followed the approach applied to temperature and analysed summer and winter precipitation sums. Although precipitation is far away from being normally distributed on short time scales, by virtue of the central limit theorem, sums over half a year are nearly normally distributed. Thus, our methods can be applied to the precipitation sums. Nevertheless, the distributions of the correlation coefficients for summer and winter precipitation are more patchy than for temperature. Low pass filtering increases the correlation coefficient locally up to 0.6 and generally better results are found in eastern Europe and the Iberian Peninsula than in the rest of Europe. However, most of the ρ's are non-significant; thus decadal predictions of total precipitation sums should be regarded carefully.

As regards fidelity, no differences exist between model and observational climatological distributions for both summer and winter.

Over most European regions the ensemble is underconfident, i.e. the ensemble spread is too large. Furthermore, the patchiness of the observed patterns hampers the understanding and interpretation of the results.

6 Conclusions and outlook

Finally, we can conclude that our analysis of the predictability of decadal variability on the regional scale is a large step to a better understanding of decadal climate processes. Several aspects of the feasability and prospects of regional decadal prediction have been investigated, but various questions still remain open. This includes e.g. the finding that the predictive skill of decadal precipitation anomalies is quite limited. This opens up the perspective and challenge to find better ways of describing the model performance on decadal precipitation in the sense of e.g. looking at finer temporal/spatial scales, other metrics or quantities like rain duration or drought indices and extreme values.

To grasp the meaning of the good correlations of CCLM and E-OBS temperatures it could e.g. be helpful to forecast special events like temperature above/below the mean. For our next work it is planned to assess the observed decadal predictability from the viewpoint of an end user. Figure 7 for instance shows the hit rate of forecasting a decade being rather a warm or cold anomaly, i.e. a binary decision of above/below the mean (trend). That means, for each grid pixel we simply counted the decades for which both the ensemble mean and the observational mean lie above/below zero. From the viewpoint of an enduser several other presentations of the forecast quality are possible, e.g. a finer partition of the forecasts into terciles or metrics based on contingency tables such as the hit rate as shown above, or scores like the Heidke skill score a.o.

Yet another question remains open, which is the potential added value of the downscaling or at least the maintenance of the forecast skill from the large scale to the regional scale. Similar to our study on decadal variability, the crucial points to investigate the added value are to choose the appropriate temporal and spatial scales of the data as well as the suitable pre-processing and metrics to tackle the problem.

Supplementary material related to this article is available online at http://www.geosci-model-dev-discuss.net/6/5711/2013/gmdd-6-5711-2013-supplement.zip.

Acknowledgements. We acknowledge the E-OBS dataset 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 program MiKlip is funded by the German ministry of education and research (BMBF).

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

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. 5716, 5718

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. 5713, 5723, 5726, 5727, 5728, 5741

Corti, S., Weisheimer, A., Palmer, T. N., Doblas-Reyes, F. J., and Magnusson, L.: Reliability of decadal predictions, Geophys. Res. Lett., 39, L21712, doi:10.1029/2012GL053354, 2012. 5716

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, L., 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. Meteor. Soc., 137, 553–597, 2011. 5715
DelSole, T. and Shukla, J.: Model fidelity versus skill in seasonal forecasting, J. Climate, 23, 4794–4806, 2010. 5719
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. 5713
Friedrich, T., Timmermann, A., Menviel, L., Elison Timm, O., Mouchet, A., and Roche, D. M.: The mechanism behind internally generated centennial-to-millennial scale climate variability in an earth system model of intermediate complexity, Geosci. Model Dev., 3, 377–389, doi:10.5194/gmd-3-377-2010, 2010. 5716
Gangstø, R., Weigel, A. P., Liniger, M. A., and Appenzeller, C.: Methodological aspects of the validation of decadal predictions, Clim. Res., 55, 2013. 5713
Giorgi, F., Jones, C., and Asrar, G.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bull., 58, 175–183, 2009. 5715
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. 5713, 5716, 5717, 5718
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. 5716
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. 5714
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. 5714
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., Rader-

GMDD
6, 5711–5745, 2013

Decadal forecast

S. Mieruch et al.

macher, 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, Reg. Environ. Change, 1–16, 2013. 5715

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. 5713

Kharin, V. V. and Zwiers, F. W.: Improved seasonal probability forecasts, J. Climate, 16, 1684–1701, 2003. 5719

Khodayar, S., Sehlinger, A., and Feldmann, H.: Sensitivity of soil moisture initialization for decadal predictions under different regional climatic condititons in Europe, Int. J. Climatol., in review, 2013. 5716

Latif, M., Delworth, T. L., Dommenget, D., Drange, H., Hazeleger, W., Hurrell, J. W., Keenlyside, N. S., Meehl, G. A., and Sutton, R. T.: 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. 5716

Lorenz, E. N.: Deterministic nonperiodic flow, J. Atmos. Sci., 20, 130–141, 1963. 5713

Mason, S. J.: From dynamical model predictions to seasonal climate forecasts, in: Seasonal Climate Variability: Forecasting and Managing Risk, Springer Netherlands, 2008. 5720

Matei, D., Pohlmann, H., Jungclaus, J., Müller, W., Haak, H., and Marotzke, J.: Two tales of initializing decadal climate prediction experiments with the echam5/mpi-om model, J. Climate, 25, 8502–8523, 2012. 5715

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. 5713, 5715, 5716

Murphy, A. H.: Skill scores based on the mean square error and their relationships to the correlation coefficient, Mon. Weather Rev., 116, 2417–2424, 1988. 5719

Palmer, T. N., Doblas-Reyes, F. J., Weisheimer, A., and Rodwell, M. J.: Toward seamless prediction: calibration of climate change projections using seasonal forecasts, B. Am. Meteorol. Soc., 89, 459–470, 2008. 5713

GMDD
6, 5711–5745, 2013
Decadal forecast
S. Mieruch et al.

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., 41, 1–24, 2013. 5714, 5715

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. 5713

Räisänen, J. and Ylhäisi, J. S.: How much should climate model output be smoothed in space?, J. Climate, 24, 867–880, 2011. 5717

Rockel, B., Will, A., and Hense, A.: The regional climate model COSMO-CLM (CCLM), Meteorol. Z., 17, 347–348, 2008. 5714

Shannon, C. E.: A mathematical theory of communication, Bell Syst. Tech. J., 27, 623–656, 1948. 5720

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., 1–14, 2012. 5713

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. 5714

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012. 5715

Thiebaux, H. J. and Zwiers, F. W.: The interpretation and estimation of effective sample size, J. Climate Appl. Meteor., 23, 800–811, 1984. 5719

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. 5713

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.,

GMDD
6, 5711–5745, 2013
Decadal forecast
S. Mieruch et al.

Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., 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. Meteor. Soc., 131, 2961–3012, 2005. 5715

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. 5713

Weigel, A. P., Liniger, M. A., and Appenzeller, C.: Seasonal ensemble forecasts: Are recalibrated single models better than multimodels?, Mon. Weather Rev., 137, 1460–1479, 2009. 5720, 5721

Wilks, D. S.: On "field significance" and the false discovery rate, J. Appl. Meteor. Climatol., 45, 1181–1189, 2006. 5719

Table 1. Number of years for the moving average filter. Example for the decade 1970 and the 5 yr moving average filter. Details see text.

Averaged point	Basis for averaging	Number of years
1971	1971–1973	3
1972	1971–1974	4
1973	1971–1975	5
1974	1972–1976	5
1975	1973–1977	5
1976	1974–1978	5
1977	1975–1979	5
1978	1976–1980	5
1979	1977–1980	4
1980	1978–1980	3

Table 2. Hit rate of a binary forecast. Details see text.

Łódź	Karlsruhe	Rome
100 %	80 %	40 %

GMDD
6, 5711–5745, 2013

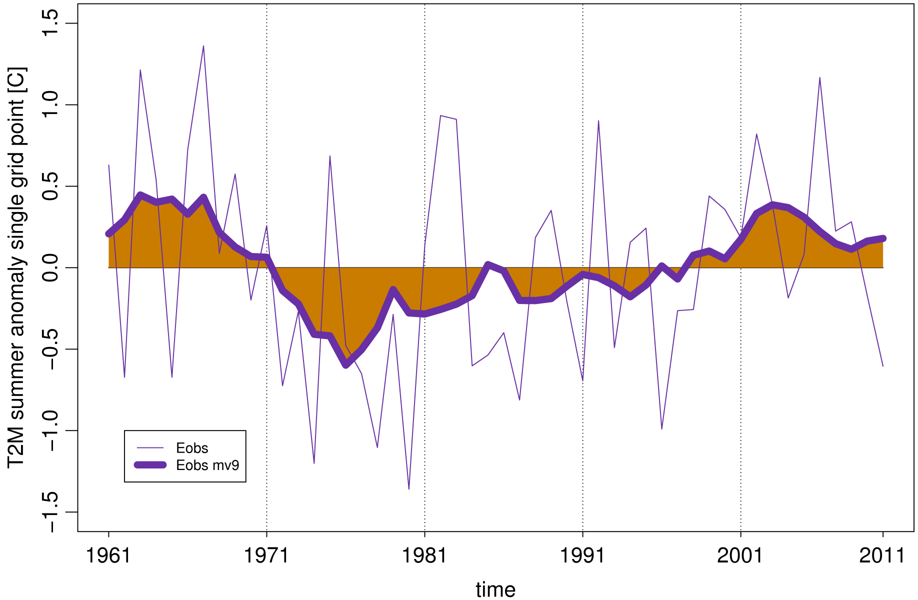

Fig. 1. Summer half-year observational (E-OBS) temperature anomalies at a single grid point near Łódź in Poland shown as a thin line. The long term mean and trend have been removed by a linear regression. Additionally a moving average filter of 9 yr is applied to extract the potential predictable decadal signal depicted as a thick line.

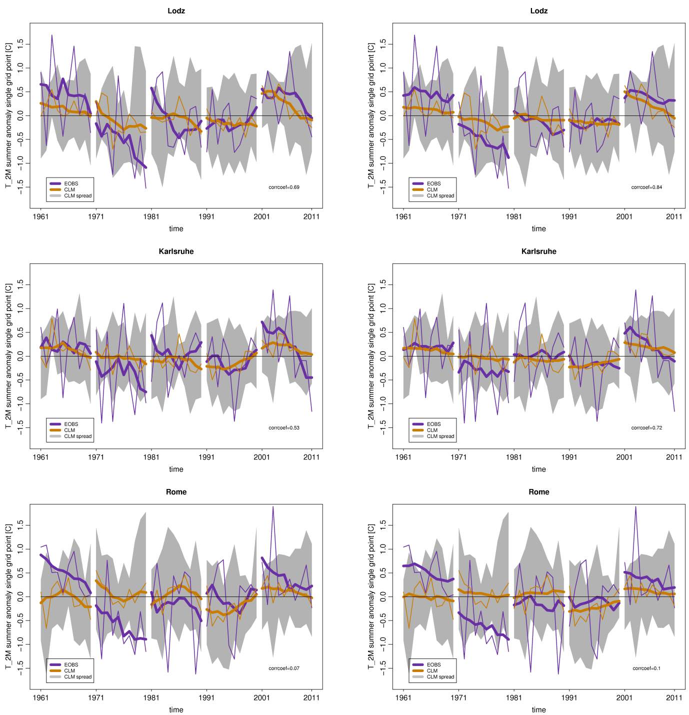

Fig. 2. Summer half-year temperatures of E-OBS (lilac) and CCLM (orange) at three different sites in Europe. The original data are shown as thin lines, whereas the 5 (left panels) and 9 (right panels) year moving average filters are plotted as thick lines. The grey shaded area depicts the CCLM ensemble spread, i.e. the standard deviation over the ensemble. Additionally, because of the decadal initialization, the decades have been separated from each other for visual inspection. Top: Time series near Łódź (Poland), high correlations have been observed of 0.69 and 0.84 for the 5 and 9 yr filters. Middle: E-OBS and CCLM at Karlsruhe (Germany) medium correlation coefficients have been estimated here of 0.53 and 0.72 for the two filters. Bottom: No decadal predictability was found near Rome (Italy) with correlation coefficients in the order of only 0.1 even for the filtered time series.

GMDD
6, 5711–5745, 2013

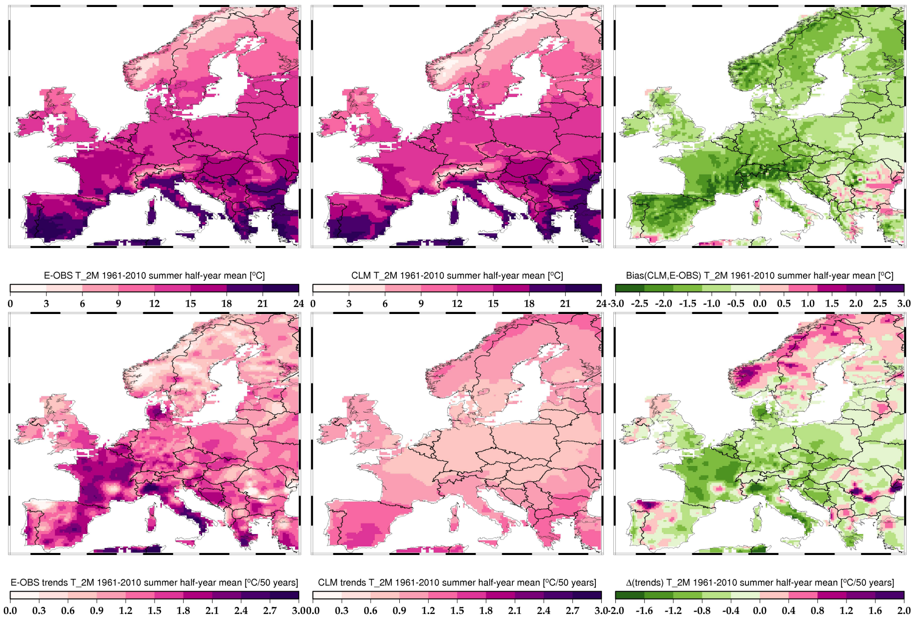

Fig. 3. Top: E-OBS and CCLM (ensemble mean) summer temperatures from 1961–2010 together with the bias. The patterns agree well, however a cold bias in the order of −2 to 0 °C exists, which however is known (Berg et al., 2013). Bottom: E-OBS and CCLM linear trends (1961–2010). The CCLM trends are generally too weak and with respect to E-OBS much smoother distributed. However, in a wide sense, the patterns agree with stronger trends in the south and weaker trends in northern Europe.

GMDD
6, 5711–5745, 2013

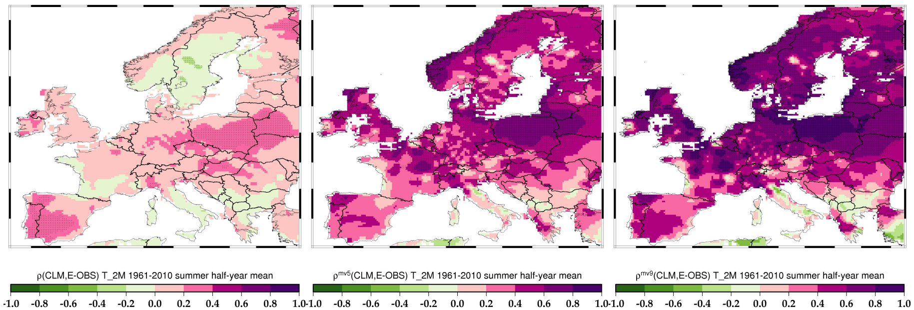

Fig. 4. Here, the summer half-year temperature correlation between E-OBS anomalies and the CCLM ensemble mean anomalies is shown. Significant correlation coefficients are represented by stippling. Left: correlation coefficients for the original unfiltered data are shown. Middle: The application of a 5 yr moving average filter is beneficial and increases the correlation coefficient. Right: A 9 yr moving average filter further increases the correlation in northern Europe, but is more or less ineffective in southern Europe.

GMDD
6, 5711–5745, 2013

Decadal forecast

S. Mieruch et al.

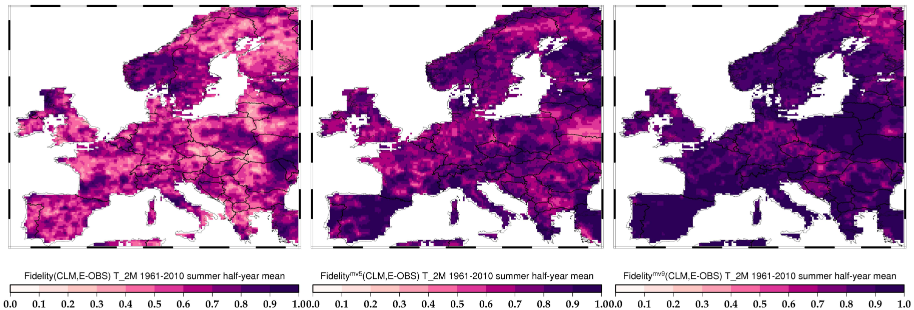

Fig. 5. To depict the fidelity, i.e. the agreement between model and observational climatological distributions, measured here with a χ^2 test, we plotted the derived p values on the European maps. A p value near zero indicates differing distributions, whilst a p value near unity supports the hypothesis of equal distributions. Shown are the results for the original data (left) and for the 5 and 9 yr filtered data (middle, right).

GMDD
6, 5711–5745, 2013

Decadal forecast

S. Mieruch et al.

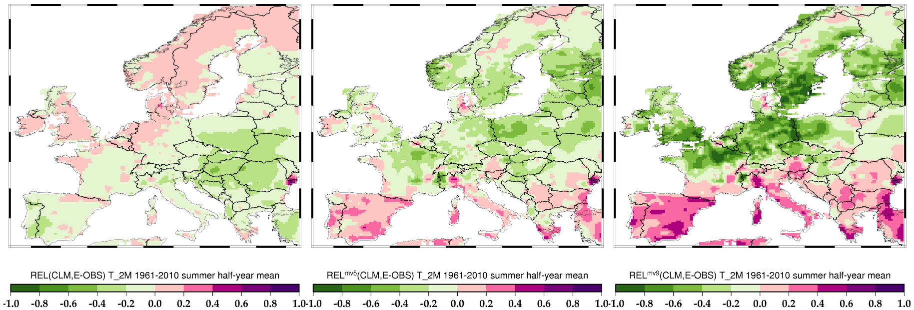

Fig. 6. The reliability of the summer temperature CCLM ensemble is a measure how well the ensemble spread covers the model error (rmse). A perfect reliability is zero, an overconfident ensemble has positive reliability and an underconfident ensemble has negative reliability. Shown are the results for the original data (left) and for the 5 and 9 yr filtered data (middle, right).

GMDD
6, 5711–5745, 2013

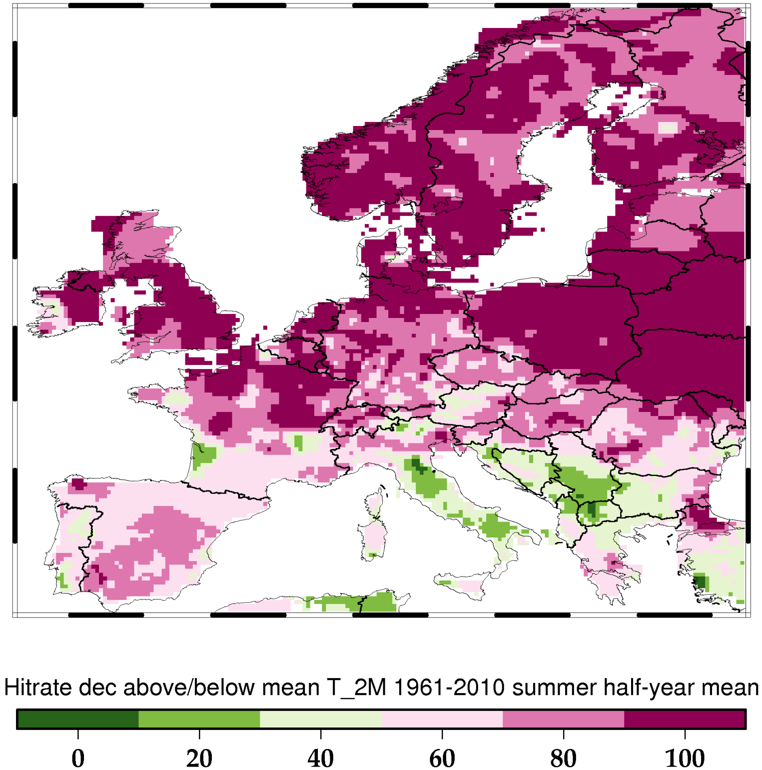

Fig. 7. Hit rate of decadal binary forecasts for summer temperature. Details see text.

GMDD
6, 5711–5745, 2013