Final Draft
of the original manuscript:

Proemmel, K.; Geyer, B.; Jones, J.M.; Widmann, M.:

Evaluation of the skill and added value of a reanalysis-driven regional simulation for Alpine temperature

In: International Journal of Climatology (2009) Wiley

DOI: 10.1002/joc.1916

Evaluation of the Skill and Added Value of a Reanalysis-driven Regional Simulation for Alpine Temperature

Short title: Evaluation of a regional simulation for Alpine temperature

Kerstin Prömmel (1 $), Beate Geyer (1), Julie M. Jones (1 *), Martin Widmann (1 #)

(1) GKSS Research Center, Max-Planck-Str. 1, 21502 Geesthacht, Germany

$ current affiliation: Institute for Meteorology, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany

* current affiliation: Department of Geography, University of Sheffield, Sheffield S10 2TN, United Kingdom

current affiliation: School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, United Kingdom

e-mail: kerstin.proemmel@met.fu-berlin.de

phone: +49 (0)30 838 71121

fax: +49 (0)30 838 71160

Abstract

To determine whether the use of regional climate models improves the representation of climate is a crucial topic in climate modelling. An improvement over coarser-scale models is expected especially in areas with complex orography or along coastlines. However, some studies have shown no clear added value for regional climate models. In this study a high-resolution regional climate simulation performed with REMO 5.0 over the whole of Europe over the period 1958-1998 is analysed for 2 m temperature over the European Alps and their surroundings called the Greater Alpine Region (GAR). The simulation is driven by perfect boundary conditions at the lateral boundaries provided by the ERA40 reanalysis and spectral nudging of the large-scale wind fields towards ERA40 values for the upper layers inside the model domain. The added value of the regional simulation (1/6° resolution) is analysed with respect to the driving reanalysis (1.125° resolution).

Both the REMO simulation and the ERA40 reanalysis are validated against different station datasets of monthly and daily mean 2 m temperature. Correlation analysis shows that the temporal variability of temperature is well represented by both REMO and ERA40, whereas both show considerable biases. The REMO bias reaches 3 K in summer in regions known to experience a problem with summer drying in a number of regional models. The comparison of REMO and ERA40 shows that an added value of the former exists for all regions in winter. For the regions surrounding the Alps the added value is absent in summer, whereas in the inner Alpine subregions with most complex orography, REMO performs better than ERA40 during the whole year. The only moderate value added by REMO in this hindcast setup may be partly explicable by the fact that meteorological measurements are assimilated in the ERA40 reanalysis but not in the REMO simulation.

1. Introduction

1.1 Hindcasts and validation

Regional Climate Models (RCMs) are an important tool for the regionalisation of global simulations for past and future climates performed with General Circulation Models (GCMs). The horizontal resolution of the GCMs is usually not finer than about 100 km, while the RCMs nowadays typically have resolutions between 10 km and 60 km (Giorgi et al., 2001 and references herein; Wang et al., 2004).

A special type of simulation for the past are global reanalyses such as the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis (Kalnay et al., 1996) or the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalyses ERA15 (Gibson et al., 1997) and ERA40 (Uppala et al., 2005), where data assimilation methods are used to find optimal estimates for past atmospheric states that are consistent with meteorological observations and the model dynamics. These global reanalyses cover the last 15 to 60 years and can be used to provide so-called perfect boundary conditions for RCMs for performing regional hindcasts. These hindcasts provide reconstructions of a large number of meteorological variables in three dimensions, which are needed in areas where no high-resolution data exist or in complex terrain where the observations are unrepresentatively distributed (Bromwich et al., 2005). Another approach to obtain estimates for past atmospheric states would be regional reanalyses using state-of-the-art data assimilation (e.g. Mesinger et al., 2006). However, such regional reanalyses are computationally demanding and are therefore in many regional climate applications replaced by hindcasts driven by global reanalyses. In addition to the performance of hindcasts another main application of RCMs is dynamical downscaling of GCM simulations of future climate change.

The validation of RCMs is mostly based on the comparison of simulations for the past with observations and the results are thus directly applicable to assessing the skill of hindcasts. Using these validation results to assess the skill of simulations for the future is not straightforward, but is nevertheless an important approach to determine the trustworthiness of models for future climate change applications (Giorgi et al., 2001; Räisänen, 2007). Before simulating future climate change with a RCM, simulations driven by perfect boundary conditions should be validated to detect systematic biases of the RCM that are primarily caused by internal model dynamics and physics (Giorgi et al. 2001). Boundary conditions for present climate derived from standard GCM simulations would already induce errors in the large-scale fields due to the fact that GCMs have limited ability to simulate present climate (Christensen et al., 1997; Giorgi et al., 2001). The validation of RCMs driven by standard GCM simulations for present climate, which is a validation of the GCM-RCM-system, is at best possible by statistically evaluating the climatologies and frequency distributions (Giorgi et al., 2001), whereas hindcasts do not only provide reconstructions of climate but are also crucial for the validation of RCMs for individual events and temporal variability. Hindcasts covering the second half of the 20th century have been analysed in many studies for pure model validation (e.g. Hagemann et al., 2002; Vidale et al., 2003: Feser, 2006; Sotillo et al., 2006; Bergant et al., 2007; Silvestri et al., 2008), as input for hydrological applications (e.g. Sotillo et al., 2005; Kotlarski et al.. 2005), or as high-resolution regional climate datasets (e.g. Bromwich et al., 2005; Sotillo et al., 2006).

A second validation approach focusing on the downscaling ability of a specific RCM setup are so called Big-Brother Experiments (Denis et al. 2002), where a RCM driven by a low-pass filtered high-resolution simulation is compared to the original high-resolution simulation, which also contains small spatial scales. Therefore, differences are attributed only to errors associated with the nesting and downscaling technique.

1.2. Added value

In both applications, regionalisation of present and future climate, it is expected that RCMs do not only yield results with just a higher resolution but that the explicit simulation of smaller-scale processes and the more detailed representation of the surface boundary conditions (orography, land-sea-contrast, land use etc.) makes the RCM simulations more realistic than the GCM results (Denis et al., 2002; Wang et al., 2004).

Although the skill of RCMs has been assessed in a number of validation studies, these studies do not explicitly address the crucial question of whether the skill of the RCM is actually better than the skill of the driving global model, in other words whether the RCM adds value to the global model. Added value is only expected on regional scales as the large scales are already well resolved by the global model (Feser, 2005). The added value analysis of hindcasts compared to the driving reanalysis shows whether the higher resolution leads to a more realistic reconstruction of present climate or whether the reanalysis, which includes assimilated observations, represents present climate more realistic despite the coarser resolution. As many variables are spatially quite homogenous and thus may be well constrained by the assimilated observations, a higher resolution does not necessarily have to lead to a better representation.

The value added by RCMs appears to have been analysed in only a few studies, which were all published recently despite many years of use of RCMs. These studies fall into two groups. In the first group the RCM skill is compared to standard GCM simulations (Duffy et al., 2006; Seth et al., 2007), while in the second group the skill of reanalysis-driven RCM hindcasts is compared to the driving reanalyses (Roads et al., 2003; Sotillo et al., 2005; Sotillo et al., 2006; Feser, 2006). These studies show that a general added value of the RCM in comparison to the driving model or reanalysis can not be found. Duffy et al. (2006) compared different RCMs over the Western United States with gridded observational data and found added value for precipitation for some, but not all RCMs, whereas for temperature they found an improvement for all RCMs in comparison to the driving GCM. Seth et al. (2007) compared precipitation of a regional and a global simulation with gridded observations over South America and concluded that the RCM did not add value to the driving GCM simulation. They suggest advancement of the physical parameterisations in both GCMs and RCMs to improve the simulations especially of tropical climate.

A comparison of four different RCMs and of the driving NCEP/NCAR reanalysis with a gridded precipitation climatology was performed by Roads et al. (2003) over South America. Over the Andes precipitation is overestimated by all models and no single RCM but the ensemble mean of all RCMs performed better than the reanalysis for some months. The authors see a reason for this low performance in the parameterisations used in RCMs which are derived from GCMs, and are not optimized for the regional scales. Sotillo et al. (2005) analysed the improvement of the RCM REMO compared to the NCEP/NCAR reanalysis at 15 offshore stations in the Mediterranean and Atlantic for 2 m temperature, mean sea level pressure and 10 m wind field. For 2 m temperature they found added value introduced by REMO mainly for extreme values. Sotillo et al. (2006) concentrated on a winter precipitation hindcast performed with REMO over the Iberian Peninsula. Based on a comparison to a high-resolution station database, they found added value not only for total amount values but also for the spatial distribution. Feser (2006) applied a spatial two-dimensional filter (Feser and von Storch, 2005) to separate the temperature and air pressure fields into large (larger than 700 km) and medium (250-550 km) spatial scales. The comparison of REMO simulations over Europe driven by the NCEP/NCAR reanalysis, as well as of the reanalysis, to gridded observations yielded added value for temperature over Europe mainly on medium scales, which has already been suggested by Laprise (2003). This improvement compared to the reanalysis was even larger if the large-scale wind field was nudged to the RCM. The positive influence of spectral nudging was already shown by Weisse and Feser (2003) for wind fields and wave heights.

1.3. Station vs. grid box

The validation of RCM simulations can be performed either against gridded data or against station data. Obviously, the advantage of using gridded data is that grid box values in GCMs and RCMs represent area means. However, gridded data can include errors that have been introduced through the interpolation method. This may be of particular concern over mountainous areas or along coastlines where the estimation of area means from station networks may be difficult due to the relatively low correlations between values at different locations (this effect is quantitatively discussed for rainday frequency in Osborne and Hulme (1997)). The estimation of absolute area means rather than anomalies is further complicated over complex orography by the strong height dependence of many meteorological variables combined with the usually non-representative height distribution of the station network. Furthermore, it should be noted that area means simulated in a regional model, even in a model with ``perfect'' physics, would differ from the real-world area means due to the differences between real and model orography. In other words, the model simulates a system that is different to the real one, and thus a comparison of simulated and real-world area means over complex orography is not a comparison of conceptually identical variables. Additionally, observations and simulations are affected by internal variability leading to the conclusion that a complete agreement is not expected (Räisänen, 2007). However, a comparison is still feasible to assess the performance of the model.

The choice between gridded or station data for validation may also depend on what the RCM simulation will be used for. If area means are of interest, they are the natural choice as a validation variable. If the estimation of local values from the RCMs without further postprocessing (statistical downscaling or Model Output Statistics) is of interest, which may often be the case, assessing the model skill by comparing the simulated area means against point observations will yield the practically relevant information. The orography-related differences between the model and the real world, as well as the height dependence of many variables are as much an issue for station-based validation as they are for the comparison of area means and should be taken into account in any analysis.

Most GCM or RCM validation studies compare simulations to gridded climatologies and reanalyses (e.g. Noguer et al., 1998; Giorgi et al., 2003; Vidale et al., 2003; Bergant et al., 2007; Jacob et al., 2007). Less work has been done on the validation against station temperature data (Kyselý, 2002; Moberg and Jones, 2004; Bromwich et al., 2005), which is partly attributable to the limited availability of good quality, high-resolution observations (Giorgi et al., 2001; Wang et al., 2004; Bergant et al., 2007). Moberg and Jones (2004) compare maximum and minimum temperatures from a RCM simulation with 50 km resolution with station data over Europe. For complex areas like the Alps with large height differences between the grid boxes and the stations, they advise against a detailed comparison as the height differences lead to a different beginning of snow melting which has a large effect on temperature variability. However, as pointed out above, the different orography in the model and the real world would also affect the comparison with area means, as systematic differences between the simulated and the real world do not get reduced through the spatial averaging. The work by Bromwich et al. (2005) shows in regions with most complex orography large negative temperature biases, which were explained by the too coarse resolution of the RCM which they suggest led to poor mixing in the lower atmosphere.

In the present study a high-resolution hindcast is analysed for Alpine temperature by a comparison with a dense, homogenised and quality controlled station dataset covering the whole GAR. As stated by Giorgi et al. (2001) RCM evaluation is limited due to a general lack of good quality high-resolution observed data, especially in regions with complex orography. Therefore, this dataset gave the opportunity for the validation as it is a dense dataset with a quality higher than any other dataset of this region including gridded datasets. Additionally, the added value of the hindcast compared to the driving reanalysis is analysed, which is of interest in the Alps where temperature is less spatially homogenous than in the surrounding regions. As shown by Sotillo et al. (2005) more added value for temperature is expected for extreme values. However, to analyse the added value for temperature extremes over the whole Alpine region an appropriate dataset covering the GAR is necessary, which does not exist. Therefore, the analysis is limited to mean temperature.

This paper is organised as follows: In section 2 the regional model and the datasets used for the validation and the validation method are described. The results are presented in section 3 separated for skill, added value and the comparison to daily data. Section 4 concludes the paper.

2. Data and method

2.1. Model data

The high-resolution simulation analysed in this study has been performed with the regional climate model REMO version 5.0 (REgional MOdel; Jacob and Podzun, 1997). The dynamical core of REMO is based on the numerical weather prediction model EM (Europa Modell) of the German Weather Service (DWD) (Majewski and Schrodin, 1994) and was further developed at the Max Planck Institute for Meteorology (MPI) and at the GKSS Research Centre. The parameterisations are taken from the ECHAM4 climate model (Roeckner et al., 1996) of the MPI. REMO is based on the primitive equations, which include the hydrostatic approximation, in a terrain-following hybrid coordinate system. The prognostic variables are surface pressure, horizontal wind components, temperature, specific humidity and cloud liquid water. The variable relevant for the present study is 2 m temperature which is not a prognostic variable but is determined from the prognostic values at the surface and the lowermost model layer taking into account the Monin-Obukhov similarity theory (e.g. Jacobson, 2005).

This study covers the period 1958 to 1998. The simulation has a high horizontal resolution of 1/6° x 1/6° (in rotated coordinates approximately 17 km) on 20 vertical levels in the troposphere and lower stratosphere. The area of interest analysed in this study, the Greater Alpine Region (GAR, 0° to 20°E and 40° to 50°N), is marked by a rectangle and consists of 66 x 90 grid points (Fig. 1).

The high-resolution simulation is driven by the global ERA40 reanalysis (hereafter ERA40; Uppala et al., 2005) through prescribing the values of the prognostic variables and of the sea surface temperature at the lateral boundaries and through forcing solely the large-scale horizontal wind field within the model domain by spectral nudging (von Storch et al., 2000) at every time step on levels above 850 hPa. By spectrally nudging the large-scale horizontal wind field, the regional model is prevented from deviating from the driving field on large spatial scales.

ERA40 covers the period from September 1957 to August 2002 and has a horizontal resolution of 1.125° x 1.125° (in Central Europe this corresponds to approx. 80 km x 125 km). For the assimilation of 2 m temperature in ERA40 only those stations located within a radius of 1000 km around the model grid point in question were used and those stations with a height difference to the model orography of more than 300 m were rejected (Simmons et al., 2004).

The resolution of REMO is 6.75 times larger than the resolution of ERA40 leading to a much more detailed representation of the Alpine orography (Fig. 2). In REMO the shape of the Alpine ridge is more realistic and single mountain ranges and large valleys are resolved, whereas in ERA40 the Alps are represented as a single low mountain.

2.2. Station data

In this study the high-resolution REMO simulation and the ERA40 reanalysis are compared to different instrumental temperature datasets over the GAR. The largest station dataset for this region is the HISTALP dataset covering the whole GAR and is described in detail in Auer et al. (2007). It consists of monthly homogenised, outlier-corrected and gap-filled records of temperature, precipitation, pressure, sunshine, cloudiness, relative humidity and vapour pressure. The acquisition of data was challenging as the stations are located in different countries and some changed their nationality, language and even their names over time. HISTALP is therefore the only multi-variable climate database for the whole GAR with such a quality and station density. However, until now this dataset exists only as monthly means. In this study temperature and cloudiness of the HISTALP dataset are used. For temperature the stations are displayed in Fig. 3a.

As the HISTALP dataset does not contain daily data, a second temperature dataset is used in this study, which consists of daily data from 59 stations from Austria and Switzerland and is named ZMdaily. The Austrian data were provided by the ZAMG (Central Institute for Meteorology and Geodynamics) and are described in Schöner et al. (2003). The Swiss data were provided by Meteoswiss (http://www.meteoschweiz.ch). 23 of the 59 stations also belong to the HISTALP temperature dataset. The daily mean temperature station dataset is also converted in this study to a monthly mean station dataset (ZMmonthly). One main difference to HISTALP is that the Austrian and Swiss stations are quality-controlled to a smaller extent and only homogenised at a national level. The ZMdaily/ZMmonthly stations are presented in Fig. 3b.

A further dataset used in this study is the CRU TS 2.0 dataset of the Climatic Research Unit (CRU), which is described in Mitchell et al. (2004). CRU TS 2.0 extends over the global land surface und covers the period 1901 to 2000. It is a gridded dataset with a resolution of 0.5° x 0.5° and includes the variables temperature, precipitation, diurnal temperature range, vapour pressure and cloud cover. In this study only temperature is used.

2.3. Method

In this study 2 m temperature is analysed, which depends strongly on altitude. As the complex orography of the Alps can not be fully captured by the reanalysis and not even by the high-resolution REMO simulation, large differences in altitude may occur between the stations and the corresponding grid boxes of REMO and ERA40. To avoid a bias due to altitude differences, an altitude correction is applied to the temperature of REMO and ERA40.

The application of a constant temperature lapse rate throughout the year might not be realistic and might lead to biases not caused by the model itself. Instead we use monthly varying lapse rates derived from the HISTALP station dataset. Seven pairs of low and high elevation stations are selected with a small horizontal distance between pairs. These pairs cover most areas of the Alpine chain and are listed in Table 1. The monthly mean temperature lapse rates of each pair are calculated from monthly values for the period 1958 to 1998 and are presented in Fig. 4. All pairs have nearly the same strong annual cycle with largest lapse rates from April to June and smallest values in December and January, meaning that the atmospheric layering is much more stable in winter than in late spring and early summer. The values of the lapse rates differ between the seven station pairs with largest differences in winter, probably reflecting the occurrence of temperature inversions. The shape of the annual cycle is in good agreement with a study by Rolland (2003), who analysed temperature lapse rates in northern Italy based on four datasets over at least 30 years. In winter his lapse rates averaged over four different regions range between 0.4 and 0.5 K/100 m compared to the area mean lapse rate of around 0.3 K/100 m in the present study. From April to August his mean temperature lapse rates are constant with values of around 0.65 K/100 m, whereas the lapse rates calculated here are constant with the same value of 0.65 K/100 m, but only from April to June. These differences might be due to the fact, that Rolland (2003) used a total of more than 600 stations and concentrated on an area around northern Italy. As in the present study the whole GAR is analysed, we use lapse rates based on stations covering large parts of the Alps instead of the lapse rates calculated by Rolland (2003). As spatial interpolation of the regional lapse rates is problematic, we use the monthly varying lapse rate averaged over the seven station pairs and represented by the thick line in Fig. 4.

For validation of the simulations different measures of skill are available. We use temporal correlation and bias in order to investigate the representation of variability and the systematic error of the simulation, respectively. To determine the reduction of error for the analysis of the added value, the root mean squared error (rmse), combining correlation and bias, is also calculated. Correlation, bias and reduction of error (defined in section 3.3) are calculated between the observed temperature at the stations and the altitude corrected temperature at the corresponding grid boxes of REMO and ERA40. The calculations are performed for each month separately over the whole simulation period of 41 years. It should be noted that the station data which we compare with REMO and ERA40 are partly assimilated in ERA40, which is not problematic because of the low horizontal small-scale variability of temperature. This is further discussed in section 3.3.

The three different station datasets can be used to answer additional questions. The comparison between the results based on ZMmonthly and those based on HISTALP are used to analyse whether the selection of the stations has an influence on the results. The effect of the temporal resolution can be analysed by comparing the results based on ZMmonthly to those based on ZMdaily.

To summarise the information at the 131 HISTALP stations and the 59 ZMmonthly/ZMdaily stations the validation results like correlation, bias and reduction of error are averaged over six subregions defined by Böhm et al. (2001). They performed a ~~based on~~ rotated principal component analysis, which is described in detail in Böhm et al. (2001), to identify regions of homogeneous temperature. The subregions are named West, East, South, Po Plain, Central Alpine Low Level (CALL) and High Level (HL) and are shown for HISTALP in Fig. 3a. Subregion HL is defined as stations with heights above 1500 m above mean sea level. As ZMdaily and ZMmonthly contain only Austrian and Swiss stations these datasets are limited to subregions West, East, CALL and HL and are presented in Fig. 3b.

3. Results

In this section the skill of REMO and ERA40 is evaluated by considering correlation and bias. Additionally, the performance of REMO and ERA40 is briefly compared. An investigation concerning the grid box/station problem, which was already outlined in the introduction, follows. Thereafter, the added value of REMO compared to ERA40 is analysed in detail. Finally, the skill from daily data is calculated and compared to the results based on monthly data.

3.1. The skill of REMO and ERA40 compared to monthly mean observed station data

3.1.1. Correlation

In Fig. 5a annual cycles of temporal correlation between the monthly mean temperature of HISTALP station data and REMO and ERA40, respectively, are presented for the six subregions. The generally very high correlations for both REMO and ERA40 indicate that the temporal variability of temperature in the GAR is represented quite well. For REMO, lowest correlations are observed in November and December and highest values in March and September averaged over all subregions. In March and September more anticyclonic general weather situations ("Großwetterlagen" based on Baur et al. (1944)) occur compared to other months, leading to calm weather easier to simulate, whereas in November more general weather situations occur, which lead to a sequence of different low pressure systems over Europe with more variable weather (Gerstengarbe et al. 1999). Subregions with generally very high correlations are West and East which are situated to the north of the Alps where orography is less complex and atmospheric circulation is less influenced by the mountains than south of the Alps in the lee of the mountains with regard to the prevailing wind direction (Barry, 1992). Here, lowest correlations are found for the subregion Po Plain during nearly the whole year. In summer this can be explained by frequently occurring instabilities and in winter by heavy fog, which may not be captured by REMO but affects temperature variability. The inner Alpine subregions CALL and HL have the most pronounced annual cycle with lowest correlations in winter. These low correlations are probably due to the differences in altitude between the grid box and the station which lead to differences in snow fall due to different temperatures. This would cause different snow cover and therefore different temperature variability between the model and observed data (Moberg and Jones, 2004). However, validation against observations of snow cover is difficult (Christensen et al., 1997).

The ERA40 correlations have a generally similar annual cycle to REMO, but slightly higher values in all subregions except the inner Alpine subregions CALL and HL with the most complex orography where REMO has higher correlations. As for REMO, ERA40 has highest correlations north of the Alps in subregions West and East and lower correlations south of the Alps, especially in the Po Plain. As the annual cycles are very similar for REMO and ERA40, the low correlations in winter in the inner Alpine subregions do not necessarily identify limitations specific to the regional model. The low correlations could either be caused individually by REMO and ERA40, or could be caused by ERA40 and be inherited by REMO. An example for the former case is the aforementioned difference between the grid box and the station altitudes.

The correlations between REMO and ERA40, respectively, and the second monthly mean station dataset ZMmonthly (Fig. 5b), which is limited to subregions West, East, CALL and HL, show similar annual cycles to the correlations with HISTALP. The only difference is that for REMO in subregions West and CALL correlations are slightly higher than for ERA40 and not lower as with HISTALP. These differences are however very small and can probably be attributed to the selection of stations.

3.1.2. Bias

The biases between the station datasets (HISTALP and ZMmonthly) and REMO and ERA40 averaged over the six subregions are presented for HISTALP in Fig. 5c and for ZMmonthly in Fig. 5d. For REMO the bias compared to both station datasets has similar annual cycles for subregions West, East, South and Po Plain with small values in winter and large positive values in summer. The largest positive bias with a value of about +3 K occurs in subregions East and Po Plain in August. In south-eastern Europe the positive summer bias is a common feature not only for REMO (Hagemann et al., 2002; Jacob et al., 2007) but also for other regional models (e.g. Christensen et al., 1997; Machenhauer et al., 1998; Noguer et al., 1998; Hagemann et al., 2001; Vidale et al., 2003; Räisänen et al., 2004; van den Hurk et al., 2005; Jacob et al., 2007). It is caused by too dry conditions over this region leading to reduced cloud cover which influences the surface energy fluxes. The abovementioned studies find different reasons for the dry conditions in different models. Christensen et al. (1997) suggest deficiencies in the surface scheme of the different analysed RCMs causing insufficient soil water and therefore an unrealistic drying-out of the soil. After van den Hurk et al. (2005) the too small depth of the hydrological soil reservoir in many RCMs including REMO plays an important role. However, Hagemann et al. (2001) found that the hydrological parameters like soil water holding capacity are well represented in HIRHAM4 and therefore are not the main cause for the summer drying problem. They found the inclusion of new land surface parameter fields (e.g. background surface albedo, vegetation cover) in the physical parameterisation to be more important. Hagemann et al. (2002) showed that problems in the general circulation in REMO and other RCMs lead to too little moisture advection into the region and hence, to the summer drying problem and the consequent influence on surface fluxes. Note that, unlike the REMO simulation analysed here, the simulation of Hagemann et al. (2002) was not forced by spectral nudging of the large-scale wind field inside the model domain. Hence, the general circulation in the simulation analysed here should be represented more realistically. This was shown by Feser (2006) by comparing pattern correlations for mean sea level pressure over Europe between observations and two REMO simulations, one with and one without spectral nudging. Despite this improvement, the warm bias over south-eastern Europe still exists. This suggests that in our simulation the main cause of the positive bias may not be a problem in the circulation but rather in the physical parameterisation of the surface scheme as proposed by most of the abovementioned studies.

In addition to the widely known positive summer temperature bias in subregion East, the REMO simulation analysed here shows a large positive summer bias in the subregion Po Plain. This was also identified in other RCMs (Christensen et al., 1997; Noguer et al., 1998; Räisänen et al., 2004; Moberg and Jones, 2004), but analysed less detailed than for south-eastern Europe. Moberg and Jones (2004) concluded by analysing minimum and maximum temperatures that the positive bias is a daytime problem due to the drying out of the soil in the model in summer, which is a similar reason as for south-eastern Europe. Typical features in the Po Plain in summer are frequently occurring instabilities during daytime (Cantù, 1977), which might not be captured by REMO possibly due to a too dry soil. This might cause too high temperatures due to the absence of the cooling effect of precipitation.

To investigate the possible causes of the bias in the simulation, simulated cloud cover is compared to observed cloud cover over the period 1958 to 1998 for each month separately. The observed values are taken from the HISTALP monthly mean cloud cover station dataset described in section 2.2 consisting of 52 stations, which are also included in the HISTALP temperature dataset. The cloud cover bias shows that the present simulation systematically underestimates observed cloud cover over the whole GAR by 0.1 to 0.5 (Fig. 6), whereas the ERA40 bias is very small during the whole year ranging from -0.17 to 0.05. The underestimation of REMO cloud cover is most pronounced in summer, which is consistent with the larger warm bias in this season, indicating a link of the temperature bias to drying. With respect to precipitation this REMO simulation has been validated against the gridded precipitation climatology of Frei and Schär (1998) in a study by Scheifinger (2006). He showed that REMO has a negative precipitation bias for both subregions Po Plain and East, which is most pronounced from July to February and also contributes to the positive summer temperature bias. This may not explain the whole temperature bias, but a more detailed examination of the physical causes of this positive summer bias is beyond the scope of this paper.

For both station datasets subregions CALL and HL have a similar annual cycle with positive biases in summer and negative biases in winter. As for CALL all REMO grid boxes are too high and for HL too low compared to the station location (not shown), the annual cycle of the bias can not be caused by the altitude correction. This annual cycle is in agreement with studies of Christensen et al. (1997), Noguer et al. (1998) and Vidale et al. (2003). They analysed different RCMs with a resolution of about 50 km over the whole of Europe compared to gridded data and showed similar results over the Alps. However, they did not distinguish between low and high elevations.

The warm bias in CALL and HL in summer is small compared to the summer biases in the other subregions and can therefore be attributed to the overall warm summer bias of REMO. In winter the cold biases in subregions CALL and HL are also quite small. For HL a potential explanation might by the slightly too low simulated cloud cover, which agrees with the result of Scheifinger (2006) showing that REMO underestimates precipitation at high elevations. This might cause too strong outgoing radiation and therefore lower temperatures. In CALL the negative winter bias might be caused by the fact that REMO does not resolve the valleys with their low elevations, which might lead to a less effective mixing of the lower atmosphere by valley winds than in the real world as suggested by Bromwich et al. (2005).

The REMO bias shows a quite consistent picture over the whole area when compared to both station datasets, with positive values in summer and small or negative values in winter. The comparison between HISTALP and ZMmonthly indicates that the temperature of ZMmonthly is in general higher than the temperature of HISTALP (not shown). As for both datasets the stations have nearly the same altitude averaged over the subregions, this difference might be due to the quality differences between the two datasets. The HISTALP dataset was subject to an extensive quality control including different homogeneity tests and homogenization methods covering the whole dataset (Auer et al. 2007) whereas ZMmonthly was subject to a homogenization over smaller regions.

For HISTALP the ERA40 bias has for all subregions a different annual cycle than the

REMO bias and the annual cycles also differ between the subregions. For ZMmonthly the annual cycle of the bias is quite different compared to HISTALP, whereas the annual cycle of the REMO bias is a robust feature for all subregions and for both HISTALP and ZMmonthly. The differences in the annual cycles of the biases of REMO and ERA40 lead to the conclusion that the strongly positive bias of REMO in summer is a clear feature of the regional model and identifies problematic regions. However, ERA40 also has large biases. In subregions South, Po Plain, CALL and HL in winter the bias compared to HISTALP reaches 2 K indicating problematic regions or processes in ERA40, but this must be caused by reasons different to those in REMO as the bias is in the opposite direction.

3.2. A comparison of the skill determined from grid box and station data

The problems arising when comparing model grid box data to station data especially over complex orography like the Alps have already been outlined in the introduction. To test whether the validation results depend on the type of data that the simulations are compared with, i.e. grid box or station data, the REMO simulation has also been compared to the gridded temperature dataset CRU TS 2.0 of the Climatic Research Unit (CRU; Mitchell et al., 2004), which has a resolution of 0.5°. As for the station data, correlation and bias are calculated over the period 1958 to 1998 for each month and each grid box separately. To compare the results to those from the HISTALP station data, only the grid boxes containing a station are selected and averaged over the six subregions. The degree of similarity between these results from CRU and those from HISTALP indicates then the potential level of the grid box versus station problem.

The correlation between REMO and the CRU data is in general very similar to the correlation between REMO and the HISTALP station data for all months and all subregions (Fig. 7a). Also for the bias a general similarity can be seen with a warm summer bias and a small bias in winter (Fig. 7b). However, compared to CRU the bias of REMO is more positive than compared to HISTALP, meaning gridded CRU temperatures are lower than HISTALP station temperatures. This discrepancy is most pronounced in subregions Po Plain and CALL. One explanation for this relates to the selection of stations used in the construction of the CRU data, i.e. if large number of mountain compared to valley stations are included this leads to a lower temperature in the grid box. However, the elevation of the stations is included in the interpolation and influences the calculation of temperature (New et al., 1999 and 2000), but in areas of complex orography where station density is lower, smaller-scale elevation dependencies are not captured leading to an inadequate inclusion of elevation in the construction of the CRU dataset. Another explanation could be that the HISTALP stations in subregions Po Plain and CALL are not representative for these areas. In the first case the differences would be an artefact of a technical problem, whereas they would reflect a real effect in the second case. However, it is beyond the scope of this paper to determine to what content these two reasons influence the results.

Nevertheless, the comparison grid box/grid box versus grid box/station shows that both comparisons have very similar results. Some larger differences occur in the Po Plain and in the inner Alpine regions with the most complex orography. Here, the biases should be interpreted with caution.

3.3. Added value of REMO compared to ERA40

The comparison of the performance of REMO and ERA40 has so far been separated into correlation and bias. The comparison to HISTALP showed that ERA40 has slightly higher correlations than REMO during the whole year, whereas the bias differences between ERA40 and REMO vary strongly during the year. Therefore, one cannot conclude directly whether ERA40 or REMO is closer to the observations. In order to analyse directly whether the higher resolution of REMO leads to an added value compared to ERA40, and to quantify the improvement, the reduction of error (RE) is calculated and shown in Fig. 8 for HISTALP and ZMmonthly. The RE is a measure of the skill of the regional model relative to the driving global reanalysis and combines the effects of correlation and bias. It is calculated by the following equation:

$$RE = 1 - \frac{rmse(REMO)}{rmse(ERA)}$$

The RE ranges from -∞ to +1. Positive values indicate an improvement of REMO compared to ERA40, zero indicates the same performance for REMO and ERA40, while negative values indicate that the agreement of REMO with station data is worse than for ERA40.

The analysis based on the HISTALP dataset shows an improvement of REMO compared to ERA40 for subregions West, East, South and Po Plain in winter and early spring and for subregions CALL and HL during the whole year (Fig. 8a). The largest positive values of RE for REMO are found for subregion HL during the whole year. Therefore, the higher resolution of the REMO simulation compared to the ERA40 reanalysis adds value especially in regions with the most complex orography. The performance of REMO compared to ERA40 is mainly influenced by the bias as the correlation differences are very small and the RE resembles strongly the difference between the REMO and ERA40 biases (not shown). This is most visible for subregions Po Plain, South and East in summer, where the performance of REMO compared to ERA40 is worst.

Compared to the results based on HISTALP the results based on the ZMmonthly dataset (Fig. 8b) show some differences in the performance of REMO. For subregion CALL negative values instead of positive values occur in spring but summer values are more positive. For subregion West the annual cycle strongly changed compared to the one based on HISTALP. These differences can be explained by the differences between HISTALP and ZMmonthly bias in these regions (Fig. 5c and d). A robust feature in the performance of REMO compared to ERA40 is the better performance of REMO in subregion HL during the whole year.

As mentioned in section 2.1, only temperature from stations whose elevation difference to the ERA40 orography is smaller than 300 m is assimilated in the ERA40 Reanalysis. Therefore, an independent validation and added value analysis is possible by comparing REMO and ERA40 to stations with a larger elevation difference. Even though the HISTALP dataset is large leading to a number of potentially independent stations, only 60 out of 131 stations match this criterion.

The RE for HISTALP calculated only with the independent stations not assimilated into ERA40 is presented in Fig. 9. Compared to Fig. 8a, where all HISTALP stations are included, the RE does not change drastically, except for subregion South where only one station is left. The other noticeable differences are the disappearance of the added value in West and Po Plain in winter and an improvement of the added value in CALL in summer. In subregion HL nothing changed, because all 12 corresponding stations were not included in the assimilation.

It was expected that REMO does not show a clear overall added value using the whole HISTALP dataset because some of the stations used for the comparison are assimilated in ERA40 leading to a better knowledge of temperature for ERA40. However, this analysis shows that the exclusion of station data used for the temperature assimilation does not have a large influence on the added value of REMO compared to ERA40. This is caused by the fact that the stations excluded from the assimilation are highly correlated to the assimilated stations nearby (not shown), as the horizontal small-scale variability of temperature is low. Thus, ERA40 has also knowledge of the temperature of the stations that are not assimilated, which makes it difficult for REMO to show an improvement over ERA40. The horizontal variability was analysed by correlating one representative station of each subregion with all other stations for each month (not shown). In summer the correlations mainly range from 0.5 to nearly 1. In winter the correlations are lower in particular for subregions CALL and HL, potentially due to local inversions.

3.4. The skill of REMO and ERA40 compared to daily station data

One may expect more added value of the high-resolution regional model on a daily timescale because of the higher small-scale variability associated with shorter temporal scales. Therefore, the monthly mean correlation, bias and RE were also calculated from daily data.

The comparison between REMO and ERA40 and the daily dataset ZMdaily only partly supports the hypothesis of more added value on daily timescales. As expected the correlations between the simulations and ZMdaily are lower than the correlations on a monthly basis for all subregions and both REMO and ERA40 (Fig. 10a). As for the comparison to ZMmonthly, REMO has slightly higher correlations than ERA40 in all subregions except East. This feature is most pronounced for HL in winter, where the ERA40 correlations are much lower than the correlations with ZMmonthly. This might be caused by the large altitude difference between the high level stations and the corresponding ERA40 grid boxes, which can lead to differences in snow fall as well as snow cover, which in turn affects temperature. The reason why this effect is stronger on daily timescales is unclear. By definition the daily bias averaged over each month is the same as the monthly bias and was discussed in section 3.1.2. The RE (Fig. 10b) shows as a measure of added value very similar results as for ZMmonthly. For subregion HL, REMO performs better than ERA40 during the whole year due to the very low correlation of ERA40. Therefore, for the high elevation stations REMO has a definite added value compared to ERA40 on the daily timescale, which is higher than that on the monthly timescale, except for summer where it is slightly lower.

4. Summary and discussion

Within the present study a high-resolution regional climate simulation has been analysed and validated for means and year-to-year variability for each month of Alpine temperature for the period 1958 to 1998. The simulation has been performed with the regional model REMO 5.0 with a resolution of 1/6° in hindcast mode driven by the ERA40 reanalysis with a resolution of 1.125°. The model domain covers the whole of Europe but the focus of this study is on the Greater Alpine Region (GAR) with its complex orography.

The high resolution of the simulation and the recent availability of a dense station dataset (Auer et al., 2007) provide the opportunity to validate the model against station data in the orographically complex area of the GAR and to summarise the results over six quite small subregions shown in Fig. 3. Such a detailed validation in complex orography has been possible for the first time. Due to the partly large elevation differences between the stations and the corresponding grid boxes, a temperature correction based on separate lapse rates for each month of the year, which are based on the station data, is applied. The application of a constant lapse rate of 0.65 K/100 m resulted in artificially high biases (not shown). Therefore, we recommend applying a more realistic monthly varying lapse rate for validation studies.

To analyse the impact of different types of data on the validation results we used station as well as gridded data for the comparison with REMO. Similarly to Moberg and Jones (2004), the present analysis shows in some inner Alpine regions large differences between the validation against gridded and against station data. In regions with a less complex orography the comparison with grid box data or with station data leads to very similar results.

The validation of REMO reveals a positive summer bias, which is most pronounced in the eastern GAR. This is consistent with earlier studies (e.g. Noguer et al., 1998; Hagemann et al., 2001; Vidale et al., 2003; Räisänen et al., 2004; Jacob et al., 2007; Christensen et al., 2007). In the inner Alps winter temperature is underestimated, which is in agreement with Christensen et al. (1997), Noguer et al. (1998) and Vidale et al. (2003). For the low elevations the negative winter bias can be explained by the poor representation of valleys in the model and therefore the missing ability to simulate the effective mixing of the lower atmosphere by valley winds as suggested by Bromwich et al. (2005). Concerning the high elevations the study by Kotlarski (2007) showed that the implementation of a glacier parameterisation scheme in REMO actually increases the existing negative winter bias compared to a simulation with standard REMO reflecting general shortcomings of REMO in the Alps.

By comparing the performance of a regional model in hindcast mode with the performance of a global reanalysis we found that in regions with the most complex orography the higher resolution of the regional model clearly improves the representation of temperature compared to the reanalysis. The added value is higher on daily than on monthly time scales. As in these inner Alpine regions less stations are assimilated into the reanalysis (see section 2.1) and temperature is horizontally less homogenous, the advantage of ERA40 due to assimilation is reduced. Additionally, in these regions small-scale processes like valley winds and cold air pockets influencing temperature are possibly better represented than in the global model and reanalyses. Another important aspect leading to the better performance in the regions with most complex orography is the application of a monthly varying lapse rate based on the station data instead of a constant lapse rate. The improvement of the added value is due to the largest elevation differences between the gridbox and the station in these regions leading to a larger influence of the more realistic altitude correction of temperature. In other regions where temperature is horizontally more homogenous, it is difficult for REMO to add value to the reanalysis, as added value is expected on small spatial scales (Feser, 2005), leading to a sometimes better representation of ERA40.

Besides our study there are only a few others that have analysed the added value in hindcasts with regional climate models relative to the driving, coarser resolution reanalysis (Roads et al., 2003; Sotillo et al., 2005; Feser, 2006). They show an added value of the RCM only for ensemble means (Roads et al., 2003), extreme values (Sotillo et al., 2005) and on medium spatial scales (Feser, 2006). Sotillo et al. (2005) found for temperature a larger added value inland demonstrating the influence of the better resolved orography, which is in agreement with the present study. When focusing on the Alps in the study of Feser (2006) the added value is less clear on medium spatial scales (250-550 km), especially in summer. A more detailed comparison to the present study is not possible due to the coarser resolution and the more general analysis over the whole of Europe. However, based on the study by Feser (2006) looking on both simulations with and without spectral nudging of the large-scale wind field, we also expect a worse added value for a simulation over the Alps without spectral nudging.

The similar performance of the regional hindcast and of the driving reanalysis in less complex terrain, despite the fact that observations are assimilated in the reanalysis but not in the regional model, and the better performance of the regional hindcast in regions with complex orography found in this study, show the considerable skill of regional models. However, improvements of RCMs and their parameterisation appear to be required to fully exploit their higher resolution and provide hindcasts that are consistently better than the reanalyses. The transferability of the results of this study to the application of the same RCM for future climate change is difficult as due to the assimilation of observations the reanalysis provides fields close to reality, which leads to a more rigorous added value analysis than that performed with a GCM-driven regional simulation for present climate. The latter analysis, which could compare climatologies and frequency distributions but could not include temporal correlations, would yield an added value that can be expected to be similar to an added value for a downscaled future climate change simulation. The added value from a reanalysis-driven hindcast might thus serve as a lower limit of the added value of the same RCM driven by a GCM simulation for future climate. This means that in regions with an added value in a hindcast an added value can also be expected in a future climate simulation with the same RCM. When the skill of a RCM hindcast rather than the added value is considered, the results are likely to be an upper limit of the performance of the RCM when driven by a standard GCM simulation.

Acknowledgements

The authors wish to thank Reinhard Böhm, Wolfgang Schöner, Nadine Salzmann, Martin Hoelzle, Fabio Monti and Michele Brunetti for providing station data and for assistance.

The ERA40 reanalysis data supplied by ECMWF and used in this study have been obtained from the CERA database. Frauke Feser is acknowledged for help with REMO and discussions on regional models and Sebastian Wagner for proof-reading and discussions. We also thank the two reviewers for their valuable comments. The present work was realised within the EU-funded project ALP-IMP (EVK-CT-2002-00148).

References

Auer I, Böhm R, Jurković A, Lipa W, Orlik A, Potzmann R, Schöner W, Ungersböck M, Matulla C, Briffa K, Jones P, Efthymiadis D, Brunetti M, Nanni T, Maugeri M, Mercalli L, Mestre O, Moisselin J-M, Begert M, Müller-Westermeier G, Kveton V, Bochnicek O, Stastny P, Lapin M, Szalai S, Szentimrey T, Cegnar T, Dolinar M, Gajić-Capka M, Zaninović K, Majstorović Z, Nieplova E. 2007. HISTALP - historical instrumental climatological surface time series of the Greater Alpine Region. *International Journal of Climatology* **27**: 17-46, DOI: 10.1002/joc.1377

Barry RG. 1992. *Mountain weather and climate, 2nd edition*. Routledge, London, 402 pp

Baur F, Hess P, Nagel H. 1944. *Kalender der Großwetterlagen Europas 1881-1939*. Bad Homburg

Bergant K, Belda M, Halenka T. 2007. Systematic errors in the simulation of European climate (1961-2000) with RegCM3 driven by NCEP/NCAR renanlysis. *International Journal of Climatology* **27**: 455-472, DOI: 10.1002/joc.1413

Böhm R, Auer I, Brunetti M, Maugeri M, Nanni T, Schöner W. 2001. Regional temperature variability in the European Alps: 1760-1998 from homogenized instrumental time series. *International Journal of Climatology* **21**: 1779-1801

Bromwich D, Bai L, Bjarnason GG. 2005. High-resolution regional climate simulations over Iceland using Polar MM5. *Monthly Weather Review* **133**: 3527-3547

Cantù V. 1977. The climate of Italy. In: Wallén CC (eds) *World survey of climatology volume 6: Climates of Central and Southern Europe*. Elsevier Scientific Publishing Company, Amsterdam, The Netherlands, pp 127-183

Christensen JH, Machenhauer B, Jones RG, Schär C, Ruti PM, Castro M, Visconti G. 1997. Validation of present-day regional climate simulations over Europe: LAM simulations with observed boundary conditions. *Climate Dynamics* **13**: 489-506

Denis B, Laprise R, Caya D, Côté J. 2002. Downscaling ability of one-way nested regional climate models: the Big-Brother Experiment, *Climate Dynamics* **18**: 627-646, DOI: 10.1007/s00382-001-0201-0

Duffy PB, Arritt RW, Coquard J, Gutowski W, Han J, Iorio J, Kim J, Leung L-R, Roads J, Zeledon E. 2006. Simulations of present and future climates in the Western United States with four nested regional climate models. *Journal of Climate* **19**: 873-895

Feser F. 2005. Spatial scale separation in regional climate modelling. GKSS Research Centre, Report No. 2005/9, 112 pp

Feser F. 2006 Enhanced detectability of added value in limited-area model results separated into different spatial scales. *Monthly Weather Review* **134**: 2180-2190

Feser F, von Storch H. 2005. A spatial two-dimensional discrete filter for limited-area-model evaluation purposes. *Monthly Weather Review* **133**: 1774-1786

Frei C, Schär C. 1998. A precipitation climatology for the Alps from high-resolution rain-gauge observations. *International Journal of Climatology* **18**: 873-900

Gerstengarbe FW, Werner PC, Rüge U. 1999. *Katalog der Großwetterlagen Europas*

(1881-1998). Potsdam, Offenbach

Gibson JK, Kållberg P, Uppala S, Nomura A, Hernandez A, Serrano E. 1997. ERA description. ECMWF ERA-15 Project Series 1: 71 pp.

Giorgi F, Hewitson B, Christensen J, Hulme M, von Storch H, Whetton P, Jones R, Mearns L, Fu C. 2001. Regional climate information - evaluation and projections. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (eds) *Climate change 2001: The scientific basis. Contribution of working group I to the third assessment report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, pp 583-638

Giorgi F, Francisco R, Pal J. 2003. Effects of a subgrid-scale topography and land use scheme on the simulation of surface climate and hydrology. Part 1: effects of temperature and water vapor disaggregation. *Journal of Hydrometeorology* **4**: 317-333

Hagemann S, Botzet M, Machenhauer B. 2001. The summer drying problem over south-eastern Europe: sensitivity of the limited area model HIRHAM4 to improvements in physical parameterization and resolution. *Physics and Chemistry of the Earth, Part B* **26**: 391-396

Hagemann S, Machenhauer B, Christensen OB, Déqué M, Jacob D, Jones R, Vidale PL. 2002. Intercomparison of water and energy budgets simulated by regional climate models applied over Europe. Max Planck Institute, Report No. 338, 45pp

Jacob D, Podzun R. 1997. Sensitivity studies with the regional climate model REMO. *Meteorology and Atmospheric Physics* **63**: 119-129

Jacob D, Bärring L, Christiansen OB, Christiansen JH, de Castro M, Déqué M, Giorgi F, Hagemann S, Hirschi M, Jones R, Kjellstöm E, Lenderink G, Rockel B, Sánchez E, Schär C, Seneviratne SI, Somot S, van Ulden A, van den Hurk B. 2007. An intercomparison of regional climate models for Europe: model performance in present-day climate. *Climate Change* **81**: 31-52, DOI: 10.1007/s10584-006-9213-4

Jacobson MZ. 2005. *Fundamentals of atmospheric modeling*. Cambridge University Press, 828 pp

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/ NCAR 40-year renalysis project. *Bulletin of the American Meteorological Society* **77**: 437-471

Kotlarski S, Block A, Böhm U, Jacob D, Keuler K, Knoche R, Rechid D, Walter A. 2005. Regional climate model simulations as input for hydrological applications: evaluation of uncertainties. *Advances in Geosciences* **5**, 119-125

Kotlarski S. 2007. A subgrid glacier parameterisation for use in regional climate modelling. Max Planck Institute, Reports on Earth System Science, Report No. 42, 180 pp

Kyselý J. 2002. Comparison of extremes in GCM-simulated, downscaled and observed central-European temperature series. *Climate Research* **20**: 211-222

Laprise R. 2003. Resolved scales and nonlinear interactions in limited-area models. *Journal of Atmospheric Science* **60**: 768-779

Machenhauer B, Windelband M, Botzet M, Christensen JH, Déqué M, Jones RG, Ruti PM, Visconti G. 1998. Validation and analysis of regional present-day climate and climate change simulations over Europe. Max Planck Institut für Meteorologie, Report No. 275

Majewski D, Schrodin P. 1994. Short description of the Europa-Modell (EM) and Deutschland-Modell (DM) of the Deutscher Wetterdienst (DWD). Quarterly Bulletin (April):1-31

Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jović D, Woollen J, Rogers E, Berbery EH, Ek MB, Fan Y, Grumbine R, Higgins W, Li H, Lin Y, Manikin G, Parrish D, Shi W. 2006. North American regional reanalysis. *Bulletin of the American Meteorological Society* **87**: 343-360

Mitchell TD, Carter TR, Jones PD, Hulme M, New M. 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100). *Tyndall Working Paper* **55**, Tyndall Centre, UEA, Norwich, UK

Moberg A, Jones PD. 2004. Regional climate model simulations of daily maximum and minimum near-surface temperatures across Europe compared with observed station data 1961-1990. *Climate Dynamics* **23**: 695-715, DOI: 10.1007/s00382-004-0464-3

New M, Hulme M, Jones PD. 1999. Representing twentieth-century space-time climate variability. Part 1: Development of a 1961-90 mean monthly terrestrial climatology. *Journal of Climate* **12**: 829-856

New M, Hulme M, Jones PD. 2000. Representing twentieth-century space-time climate variability. Part 2: Development of 1901-96 monthly grids of terrestrial surface climate. *Journal of Climate* **13**: 2217-2238

Noguer M, Jones R, Murphy J. 1998. Sources of systematic errors in the climatology of a regional climate model over Europe. *Climate Dynamics* **14**: 691-712

Osborn TJ, Hulme M. 1997. Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *Journal of Climate* **10**: 1885-1908

Räisänen J, Hansen U, Ullerstig A, Döscher R, Graham LP, Jones C, Meier HEM, Samuelsson P, Willén U. 2004. European climate in the late twenty-first century: regional simulations with two driving global models and two forcing scenarios. *Climate*

Dynamics **22**: 13-31, DOI: 10.1007/s00382-003-0365-x

Räisänen, J. 2007. How reliable are climate models? *Tellus* **59A**: 2-29

Roads J, Chen S, Cocke S, Druyan L, Fulakeza M, LaRow T, Lonergan P, Qian J-H, Zebiak S. 2003. International Research Institute/Applied Research Centers (IRI/ARCs) regional model intercomparison over South America. *Journal of Geophysical Research* **108**: D14, 4425, DOI: 10.1029/2002JD003201

Roeckner E, Arpe K, Bengtsson L, Christoph M, Claussen M, Dümenil L, Esch M, Giorgetta M, Schlese U, Schulzweida U. 1996. The atmospheric general circulation model ECHAM4: model description and simulation of present-day climate. Max Planck Institut für Meteorologie, Report No. 218

Rolland C. 2003. Spatial and seasonal variations of air temperature lapse rates in Alpine regions. *Journal of Climate* **16**: 1032-1046

Scheifinger H. 2006. Climatological evaluation of the REMO (regional model) precipitation simulation 1971-1999. Internal ALP-IMP project report

Schöner W, Auer I, Böhm R, Thaler S. 2003. Qualitätskontrolle und statistische Eigenschaften ausgewählter Klimaparameter auf Tageswertbasis im Hinblick auf Extremwertanalysen. StartClim-Startprojekt Klimaschutz: Erste Analysen extremer Wetterereignisse und ihrer Auswirkungen in Österreich. Endbericht Projekt 1: 1-39

Seth A, Rauscher SA, Camargo SJ, Qian J-H, Pal JS. 2007. RegCM3 regional climatologies for South America using reanalysis and ECHAM global model driving fields. *Climate Dynamics* **28**: 461-480, DOI: 10.1007/s00382-006-0191-z

Silvestri G, Vera C, Jacob D, Pfeier S, Teichmann C. 2008. A high-resolution 43-year atmospheric hindcast for South America generated with the MPI regional model.

Climate Dynamics, DOI: 10.1007/s00382-008-0423-5

Simmons AJ, Jones PD, da Costa Bechtold V, Beljaars ACM, Kållberg PW, Saarinen S, Uppala SM, Viterbo P, Wedi N. 2004. Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *Journal of Geophysical Research* **109**: D24115, DOI: 10.1029/2004JD005306

Sotillo MG, Ratsimandresy AW, Carretero JC, Bentamy A, Valero F, González-Rouco F. 2005. A high-resolution 44-year atmospheric hindcast for the Mediterranean Basin: contribution to the regional improvement of global reanalysis. *Climate Dynamics* **25**: 219-236, DOI: 10.1007/s00328-005-0030-7

Sotillo MG, Martín ML, Valero F, Luna MY. 2006. Validation of a homogeneous 41-year (1961-2001) winter precipitation hindcast dataset over the Iberian Peninsula: assessment of the regional improvement of global reanalysis. *Climate Dynamics* **27**: 627-645, DOI: 10.1007/s00382-006-0155-3

Uppala SM, Kållberg PW, Simmons AJ, Andrae U, da Costa Bechtold V, Fiorino M, Gison JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Moscrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 reanalysis. *Quarterly Journal of the Royal Meteorological Society* **131**: 2961-3012

van den Hurk B, Hirschi M, Schär C, Lenderink G, van Meijgaard E, van Ulden A, Rockel B, Hagemann S, Graham P, Kjellström E, Jones R. 2005. Soil control on runoff response to climate change in regional climate model simulations. *Journal of Climate* **18**: 3536-3551

Vidale PL, Lüthi D, Frei C, Seneviratne SI, Schär C. 2003. Predictability and uncertainty in a regional climate model. *Journal of Geophysical Research* **108**: D18, 4586, DOI: 10.1029/ 2002JD002810

von Storch H, Langenberg H, Feser F. 2000. A spectral nudging technique for dynamical downscaling purposes. *Monthly Weather Review* **128**: 3664-3673

Wang Y, Leung LR, McGregor JL, Lee D-K, Wang W-C, Ding Y, Kimura F. 2004. Regional climate modeling: progress, challenges, and prospects. *Journal of the Meteorological Society of Japan* **82**: 1599-1628

Weisse R, Feser F. 2003. Evaluation of a method to reduce uncertainty in wind hindcasts performed with regional atmosphere models. *Coastal Engineering* **48**: 211-225

station names	abbr.	elev. difference	horiz. distance	yearly mean lapse rate	subregion
Innsbruck/Patscherkofel	INN/PAK	1638 m	9.00 km	0.54K/100m	CALL
Zell am See/Schmittenhöhe	ZEL/SCH	1207 m	3.91 km	0.38K/100m	CALL
Bad Ischl/Feuerkogel	BIL/FEU	1149 m	13.00 km	0.42K/100m	East
Badgastein/Sonnblick	BGA/SON	2005 m	15.71 km	0.55K/100m	CALL
Interlaken/Jungfraujoch	ITL/JFJ	2996 m	15.78 km	0.53K/100m	West
Aosta/Gr. St. Bernhard	AOS/GSB	1928 m	18.15 km	0.61K/100m	Po Plain
Graz/Schökl	GRA/SCK	1059 m	13.44 km	0.50K/100m	East

Table 1: The seven stations pairs used for the calculation of the monthly varying lapse rate with their abbreviations, elevation differences, horizontal distances, yearly mean lapse rates and subregion of the lower station.

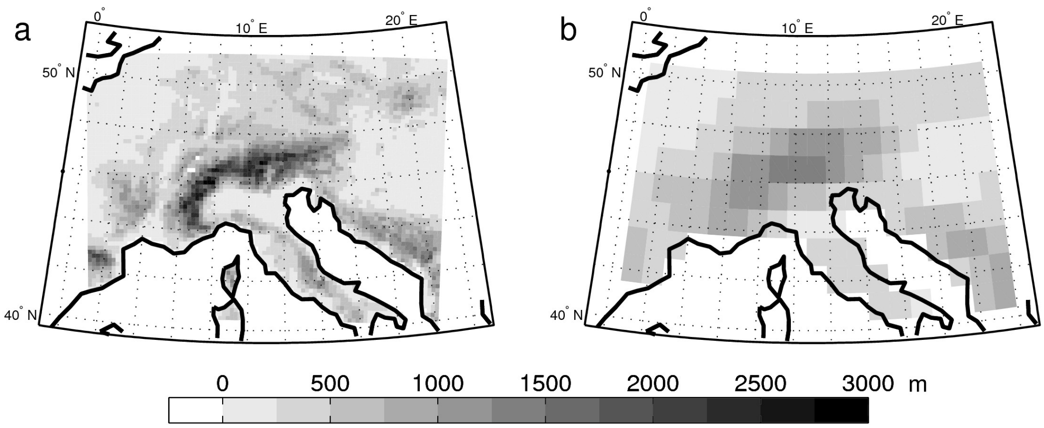

Figure 1: Whole simulation area with model orography in m. The rectangle shows the study area, Greater Alpine Region (GAR).

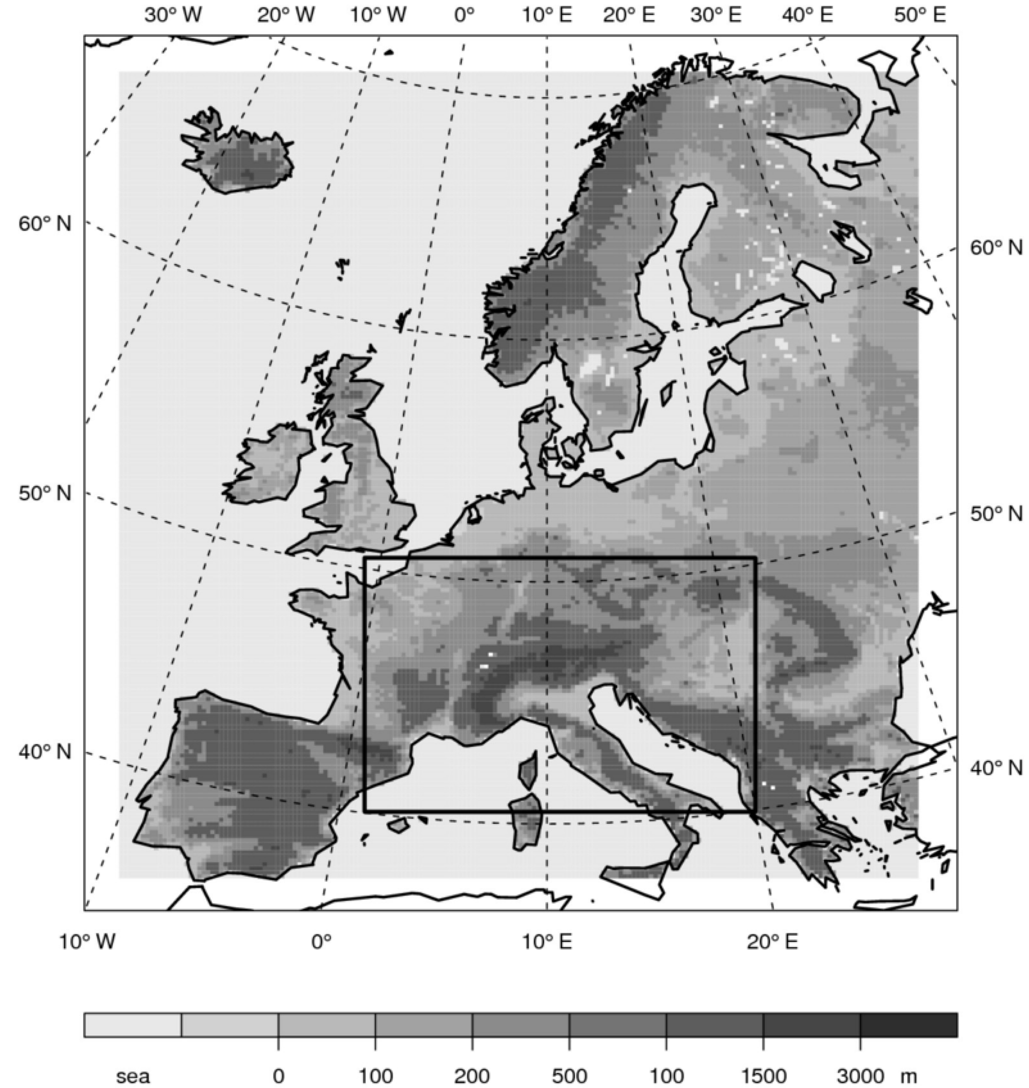

Figure 2: Orography in m of REMO (a) and ERA40 (b) in the study area.

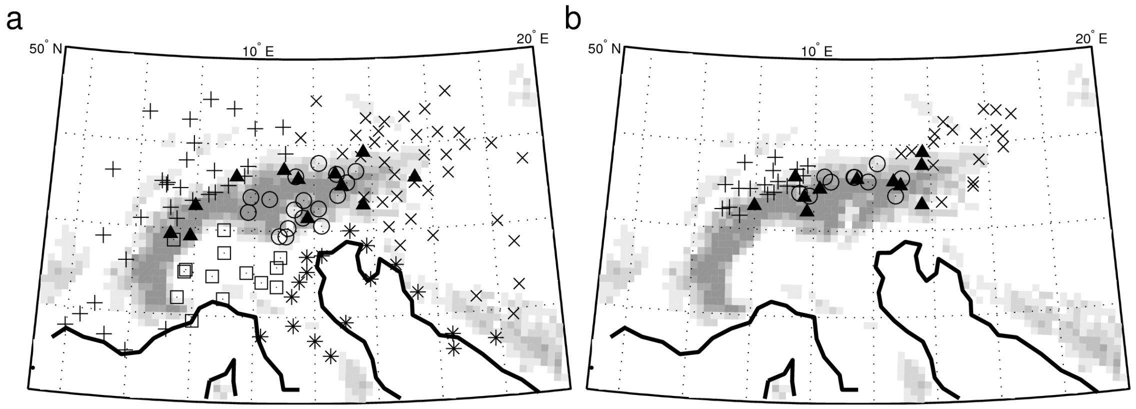

Figure 3: HISTALP (a) stations and ZMmonthly/ZMdaily (b) stations divided into subregions West (+), East (×), South (*), Po Plain (□), Central Alpine Low Level (CALL,○) and High Level(HL,▲). Shading in the background represents orography used in REMO with an interval of 500 m.

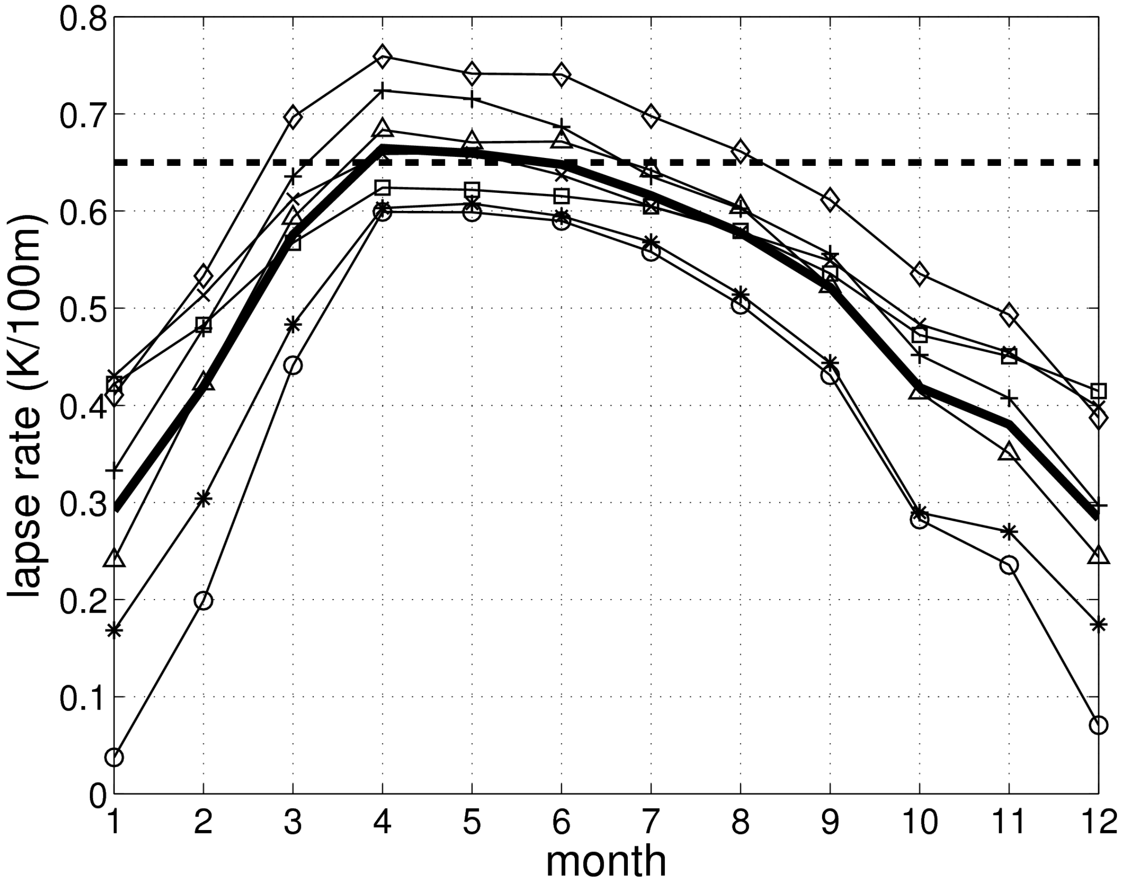

Figure 4: Annual cycles of lapse rates (K/100m) based on seven station pairs (thin lines) and the mean of all seven lapse rates (thick line). The constant lapse rate of 0.65 K/100m is indicated by the dashed thick line. Seven station pairs: INN/PAK (+), ZEL/SCH (○), BIL/FEU (*), BGA/SON (×), ITL/JFJ (□), AOS/GSB (◊), GRA/SCK (Δ).

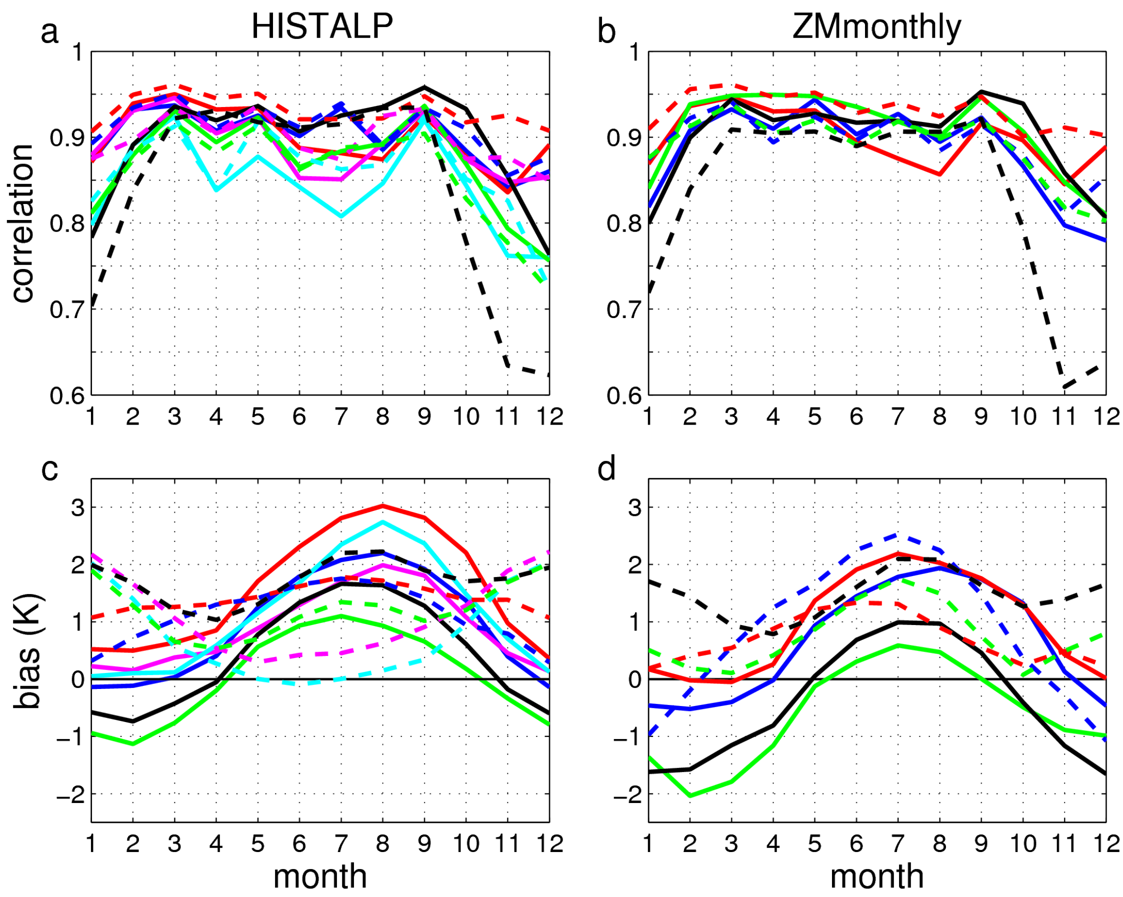

Figure 5: The annual cycles of the performance of REMO (solid) and ERA40 (dashed) are shown for the temperature correlation with HISTALP (a) and with ZMmonthly (b) and the temperature bias compared to HISTALP (c) and to ZMmonthly (d) averaged over the subregions West (blue), East (red), South (magenta), Po Plain (cyan), CALL (green) and HL (black).

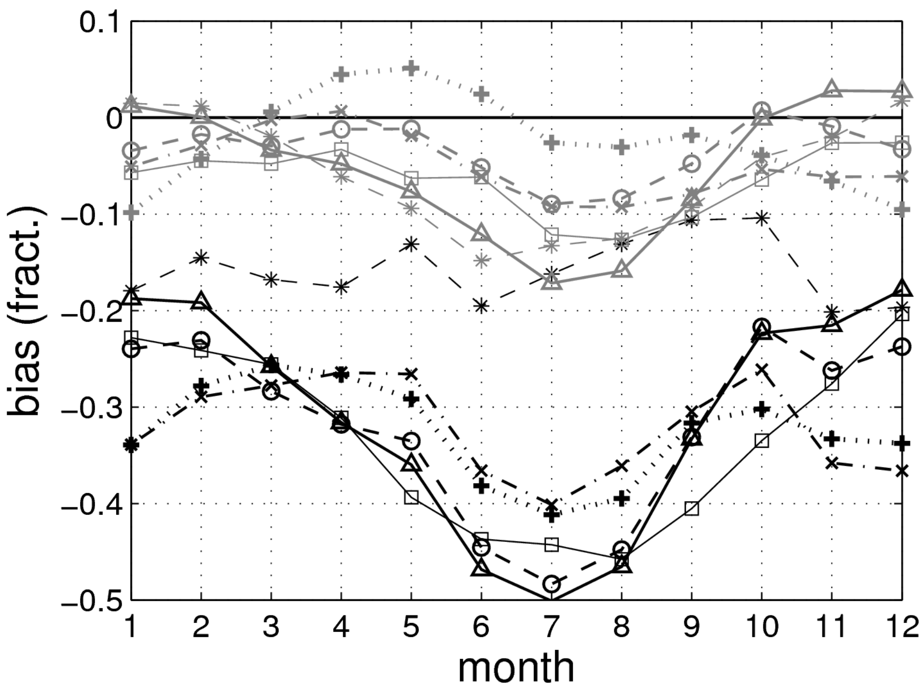

Figure 6: Annual cycles of cloud cover bias between REMO (black) and ERA40 (grey), respectively, and the HISTALP monthly mean cloud cover station dataset averaged over the subregions West (+), East (×), South (*), Po Plain (□), CALL (○) and HL (Δ).

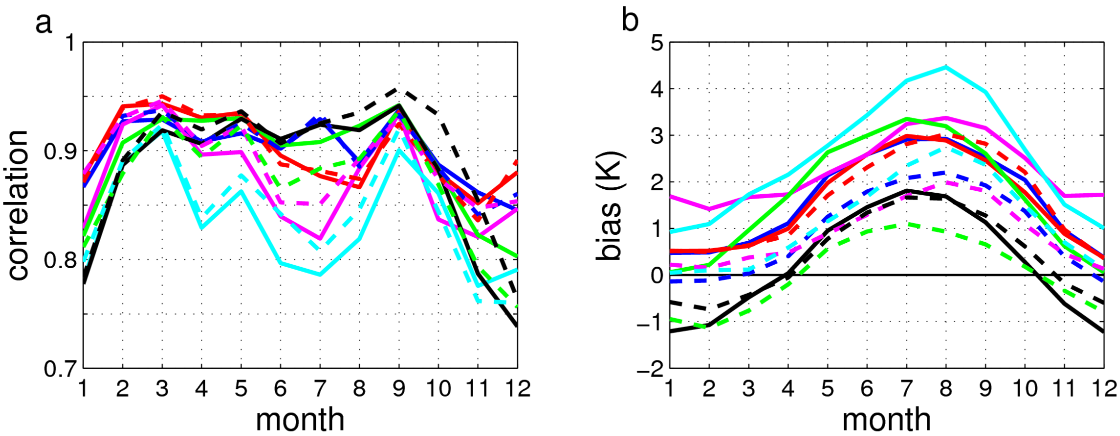

Figure 7: Annual cycles of the performance of REMO compared to the gridded CRU data (solid) and compared to the HISTALP stations (dashed) for temperature correlation (a) and bias (b) averaged over the subregions West (blue), East (red), South (magenta), Po Plain (cyan), CALL (green) and HL (black).

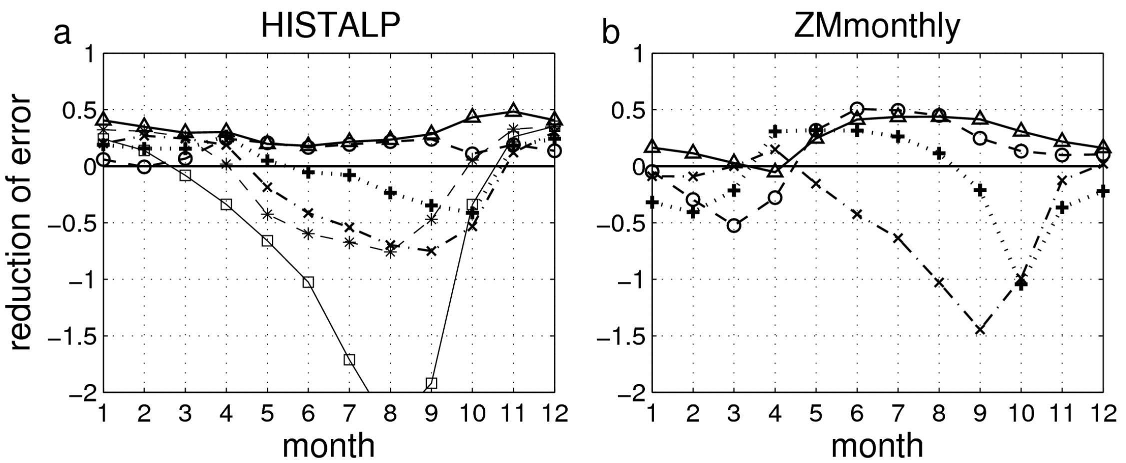

Figure 8: Annual cycles of the reduction of error of REMO temperature compared to ERA40 temperature for HISTALP (a) and ZMmonthly (b) averaged over the subregions West (+), East (×), South (*), Po Plain (□), CALL (○) and HL (Δ).

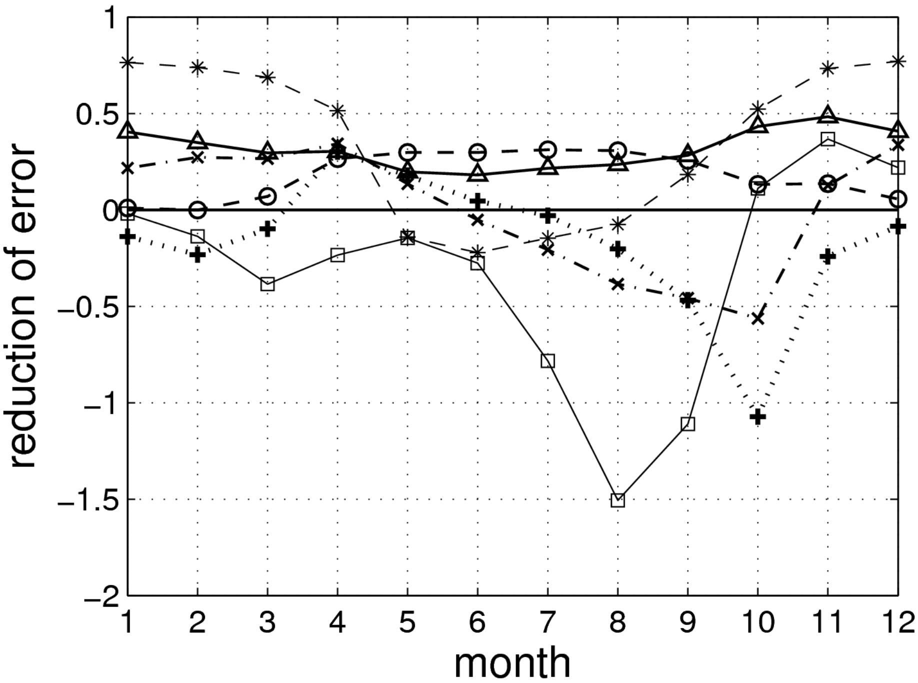

Figure 9: Annual cycles of the reduction of error of temperature calculated with the HISTALP stations not assimilated in the ERA40 reanalysis averaged over the subregions West (+), East (×), South (*), Po Plain (□), CALL (○) and HL (Δ).

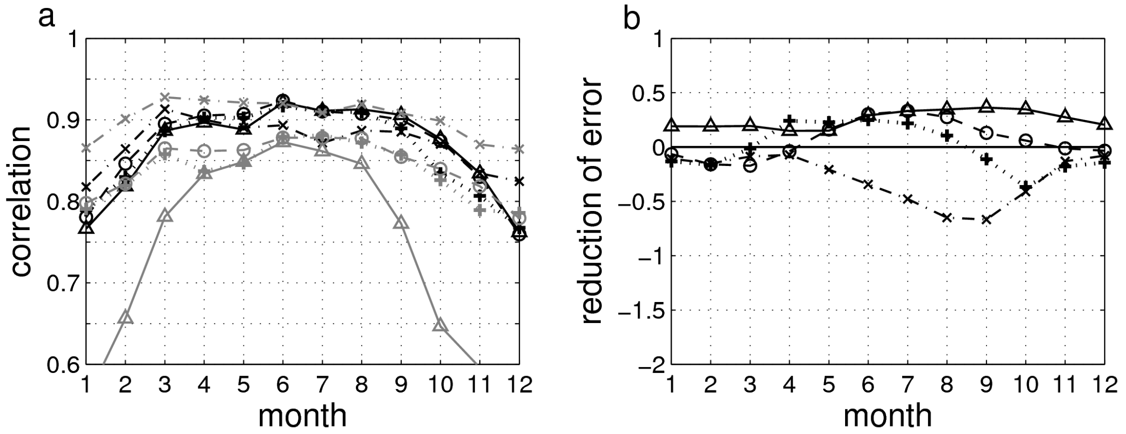

Figure 10: Annual cycles of the daily temperature correlation (a) between ZMdaily and both REMO (black) and ERA40 (grey) and of the daily reduction of error (b) of REMO temperature compared to ERA40 temperature based on ZMdaily averaged over the subregions West (+), East (×), CALL (○) and HL (Δ).