Project No. 037005

CECILIA

Central and Eastern Europe Climate Change Impact and Vulnerability Assessment

Specific targeted research project

1.1.6.3.I.3.2: Climate change impacts in central-eastern Europe

D3.2 RCM output localization methods

Due date of deliverable: 1st December 2007
Actual submission date: 30th December 2009

Start date of project: 1st June 2006   Duration: 43 months

Lead contractor for this deliverable: BOKU

Final revision

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of contents

Introduction

The climate models, although based on a set of equations and parameterizations arising from the physical reality of the climate system, often suffer from significant systematic errors. In order to use the outputs of GCMs/RCMs for consequent analyses (such as studies dealing with the impacts of climate change), additional postprocessing is desirable.

This report summarizes the effort of various CECILIA teams concentrated on different strategies of statistical correction of raw RCM data reducing the biases of model simulations and bringing a better match between the PDFs of simulated and observed data. Furthermore, approaches allowing statistical localization of the model-produced fields to subgrid resolutions were investigated, i.e. methods increasing the horizontal resolution beyond the 10 km step of regional climate models employed in CECILIA. It is demonstrated that the corrections can reduce or eliminate some of the discrepancies the models suffer from, and that additional fine details can be introduced to the RCM-simulated fields of climate variables. The material presented is divided into chapters containing description of activities done by individual partners involved in the deliverable:

BOKU (page 5), focusing on a quantile-matching corrective procedure for daily temperature, precipitation and relative humidity, simulated in the Central European region,

CUNI (page 16), testing a low-parametric alternative to the postprocessing procedure employed at BOKU and applying an elevation-based technique of localization of daily temperature and precipitation, simulated in the Central European region,

NIMH (page 28), using transformation based on the topography characteristics (elevation and gradients) before bilinear interpolation in order to minimize the interpolation error, for monthly temperature and precipitation in the Bulgarian region.

Correction of RegCM3 model output data using a rank matching approach applied on various meteorological parameters (BOKU)

Summary

A method for a post-processing model bias correction is investigated using daily data obtained from the regional climate model RegCM3 driven by reanalysis data. The evaluation of the performance is done by comparing the original model data and the corrected model data with observational data.
The regional climate model data from RegCM3 used here is a high-resolution climate simulation run with grid spacing of 10 km for the time period 1961-2000 covering the central and eastern part of Europe.

The correction method is based on using the differences of the empirical cumulative density functions of model and observation and it is applied to the model data such that the statistics of the observations are retained. The method uses correction factors that correct the model data depending on the CDF values. The correction factors are defined either as an additive correction, which was used e.g. for the temperature data or as a multiplicative correction which was used e.g. for precipitation. The correction factors are calculated on a monthly basis.
A split sample test was applied to evaluate the method. The data of the period 1961-1985 was used to calculate the correction factors and these were used then to correct the data of 1986-2000. This corrected data was then verified against the observations of this period.
The method is simple to implement and it shows very good improvement of the model data for the main statistical parameters for temperature (maximum, minimum, mean), relative humidity and precipitation although the mean absolute error is reduced only very little, because the temporal correlation is not improved. Statistical indices related to the CDF like mean and variance are improved. The correction method for temperatures and relative humidity performs very well and can be used for model output correction without any limitation. For precipitation a number of problems occur due to the distribution of precipitation improving not all aspects of precipitation statistics.

Methodology

The statistical properties of a dataset are entirely described by the probability density function (PDF) or by the cumulative density function (CDF).
This means that if two data sets have the same PDF (CDF) then they are statistically identical i.e. they have the same mean the same variance and all the other moments are also identical.

If we want to correct a dataset such that the statistical properties are recovered this could be done by correcting the data in such a way that the CDFs become identical. Therefore we construct the correction of the data using the CDFs. In fact we are not using the CDFs directly but the quantile function, which is the inverse of the CDF.

Here we investigate a method that uses the CDF of an observation dataset und the CDF of a model to find correction factors that modify the model output data in such a way that the CDFs of the observations and the corrected data become identically. Once the correction factors are found it is possible to correct all data of the model assuming that the model bias statistics don't change.

If we have data from a model and observational data then the observational data can be used to correct the model data. This can be done in various ways. One possibility is to use the CDF of both datasets and to adjust them.
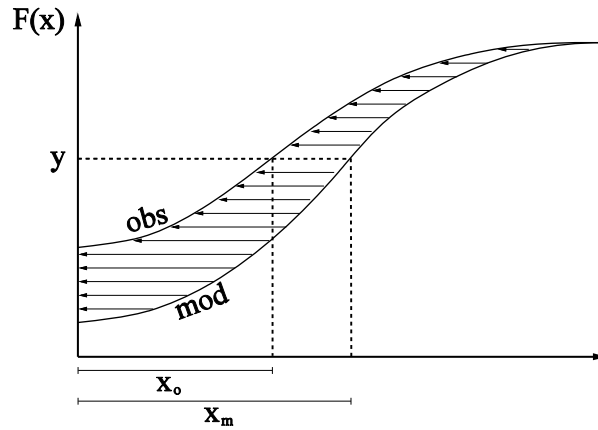

Figure 1: Illustration of the rank matching method. Shown are a CDF of observations and a CDF of model output and the illustration of correcting at the probability value **y**

To explain the correction method now in detail we look at Figure 1, which shows the two CDFs. The x-axis values can be seen as p-quantiles and the y values are probabilities. The notation **obs** and **mod** denote the CDF of the observation and the CDF of the model, respectively. Since the model CDF does not match the obs CDF very well the model data has some kind of bias or error that we want to correct. The general goal now is to correct the original model data set in such a way that the CDF of the corrected model data becomes identical with the CDF of the observations. The small arrows in the Figure between the two CDFs indicate how the CDF values have to be shifted to achieve this goal. All we do therefore is to apply a correction to the model data depending on the CDF values.

For example we want to define a correction for the probability **p**. The dotted horizontal line shows the two according **p**-quantiles. The observation p-quantile is x_o and the model p-quantile is denoted as x_m. We see that the model p-quantile is larger than the observation p-quantile. We now define two different correction factors to correct the original model data having the CDF value **p**. The first correction is an additive correction defined as:

$$f_a(y) = F_o^{-1}(y) - F_m^{-1}(y) = x_o - x_m$$

The second correction is a multiplicative correction factor written as:

$$f_m(y) = \frac{F_o^{-1}(y)}{F_m^{-1}(y)} = \frac{x_o}{x_m}$$

If we now correct a value of the original model data set that has the CDF value **p** we see that the corrected value **x$_c$**: is the observation value:

$$x_c = x_m + f_a = x_m + x_o - x_m = x_o$$

and analog for the multiplication factor. So we have corrected the model output value **x$_m$** to the observation value **x$_o$** both having the CDF value **p**. This is now done for all **p** values und the corrected model data set has then the same CDF like the observations.

A simple example shown in Figure 2 demonstrates the method showing the PDF and the CDFs of observation and model. This example was used to test the method on synthetic data showing the ability to reconstruct the PDF and the CDF using the here defined correction factors on the data. We see very clearly that the distributions of the observations and of the corrected data are very similar which is the central goal of this method.

Since this method can be used on any model and observational data set, it has to be decided critically, if it makes sense to correct the data in such a way since the method might change the original model dataset. It is rather necessary that the model data and the observation data are somehow similar and the corrections applied are relatively small. Therefore a close look at the correction factors gives insight on the changes applied.
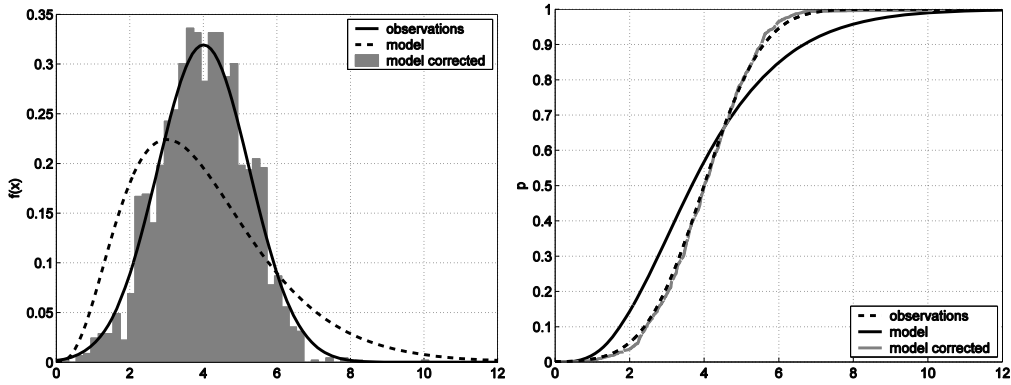

Figure 2: The left panel shows the PDFs of the synthetic data. The solid line is the PDF of the synthetic observational data. The corrected model data are shown in light grey. The synthetic model data is the dashed line which shows a shift to lower values and a gamma skewed distribution. The right side shows the corresponding CDF.

We use an empirical CDF as basis for the correction. Empirical CDFs can always be estimated by simply calculating the percentiles for a number of values, which makes this method also very useful since there is no assumption needed concerning an underlying probability function.

In Figure 3 the spatial distribution of the mean correction factors for precipitation (left side) and temperature maximum (right side) for values higher than the median are shown. The

spatial pattern of the correction factors vary between the months and parameter, but often the structure is related to the topography. For precipitation (left side) the multiplicative factors lower than 1 indicate an overestimation of the precipitation in the model and for temperature maximum the positive values indicate an underestimation of the model.
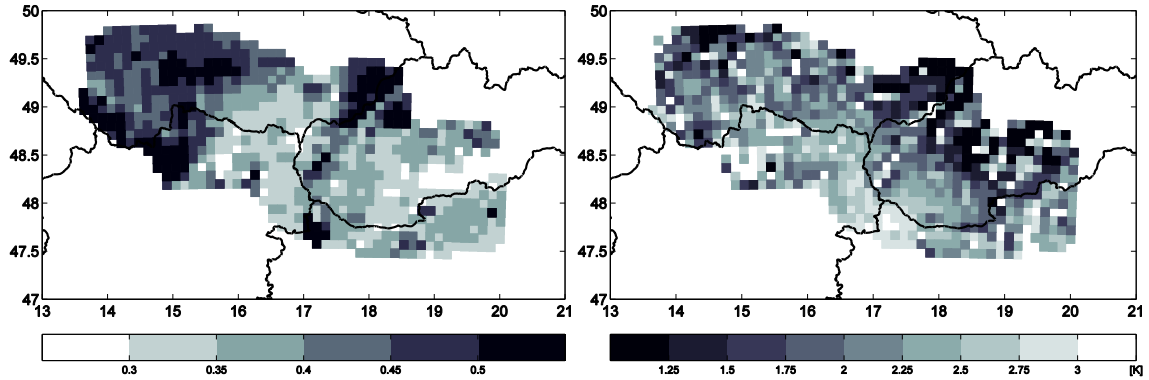

Figure 3: The spatial distribution of mean correction factors for values between **p**=0.5 and **p**=1 for RR (left) and Tmax (right)

Figure 4 shows the temporal correlation of the monthly sums of precipitation (left side) and the improvement achieved with the bias correction. This type of bias correction has no significant effect on the temporal correlation. In most regions there is no change or light improvements, but in several regions even a decrease in temporal correlation can be seen.
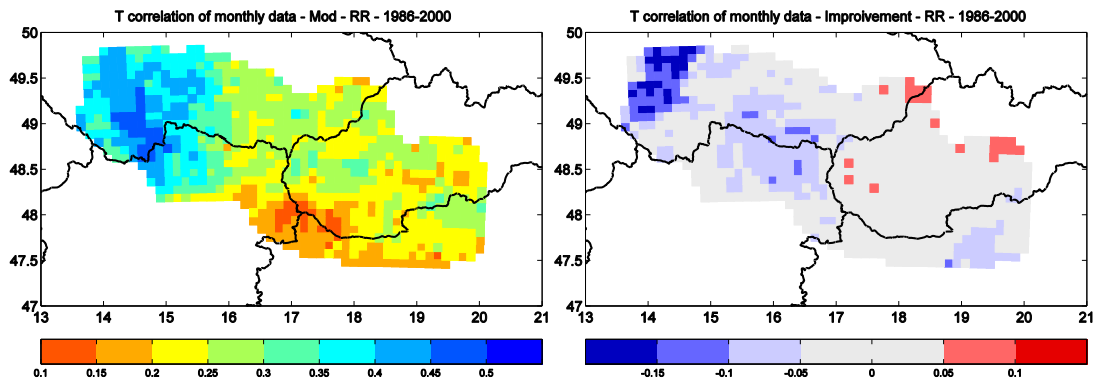

Figure 4: The left panel shows the time correlation in terms of R^2 of monthly precipitation comparing the model precipitation with the observations. The right panel shows the improvement of this time correlation using the correction method.

Evaluation and discussion

The empirical bias correction method is tested on model data and observational data. The regional climate model is forced with ERA40 reanalyses data, which allows a direct comparison of model result with observations. The calibration period where the correction factors are calculated is 1961-1985 and the evaluation period where the model data is corrected with the calculated correction factors is 1986-2000. The investigated parameters are maximum and minimum temperature, relative humidity, radiation and precipitation although we will present here only results for maximum temperature and precipitation. These two parameters are chosen to demonstrate the additive (maximum temperature) and the multiplicative (precipitation) correction as already discussed earlier.

The correction method is evaluated comparing the original model data and the corrected model data with the observations for the time period 1986-2000. These 15 years are not very long but we decided to use a longer time period (25 years) for calibration to receive more accurate correction functions.
The calibration is performed by dividing the range of the inverse CDF function into 100 sections and by calculating the midpoint values of the observations and of the model. The midpoint was chosen to use somewhat like an average value that is valid for the interval. For example the correction factor for the values between the 99th and the 100th quantile of the data are corrected with the value of 99.5 quantile.

RegCM3 produces a large number of days with very small amounts of precipitation, which are usually set to zero. To demonstrate the ability of the correction method this was not done. Rather the original model output data was used.

In order to show some detailed results two single grid points are chosen: one in the Northwest in a rather orographic complex terrain and one in the Southeast in a very low and flat part of the domain. These two points will be denoted with NW and SE respectively.

The first step of the correction process is to calculate the correction factors using data from the calibration period 1961-1985. The results for maximum temperature on a monthly basis are shown in Figure 5 and the correction factors of precipitation are shown in Figure 5. The solid line shows the correction factors for grid point NW and the dashed line for grid point SE. The grey lines are the correction factors for all other points. The panels are arranged such that line one to four show DJF, MAM, JJA and SON respectively.

We see that the correction factors of Tmax for our two selected grid points reduce the values of the model output in December and January for almost all **p** values except of values close to the maximum. From March on the character of the correction changes. The values of the original model data are then too low and have to be corrected up to 4 °K e.g. if we look at month 5 we see that for almost all **p** values the correction is more than 2 °K for grid point SE while grid point NW has up to a **p** value of 0.5 almost no correction and then increasing values up to 3.5 °K. This demonstrates the different characteristics of different grid points, which have to be taken into account for a proper correction.

The correction factors of precipitation show that **p** values up to about 0.3 are zero which means that the precipitation is set to 0 i.e. the correction method produces dry days. The maximum **p** value being 0 is the fraction of dry days as obtained from the calibration i.e. the

fraction of dry days in the observations. Since the model always has less dry days than the observations a large part of the CDF range is reduced.
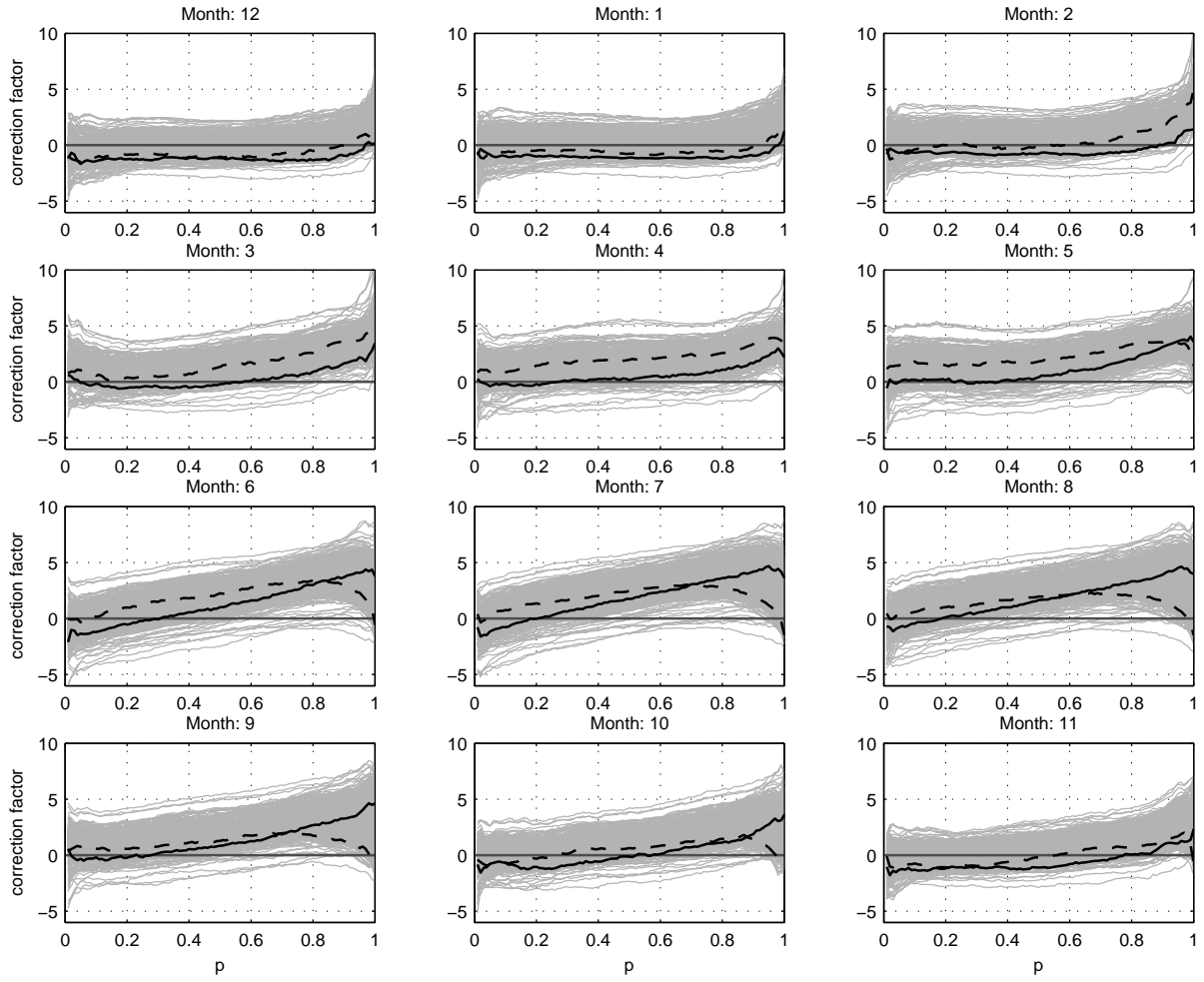

Figure 5: The monthly correction factors for maximum temperature.
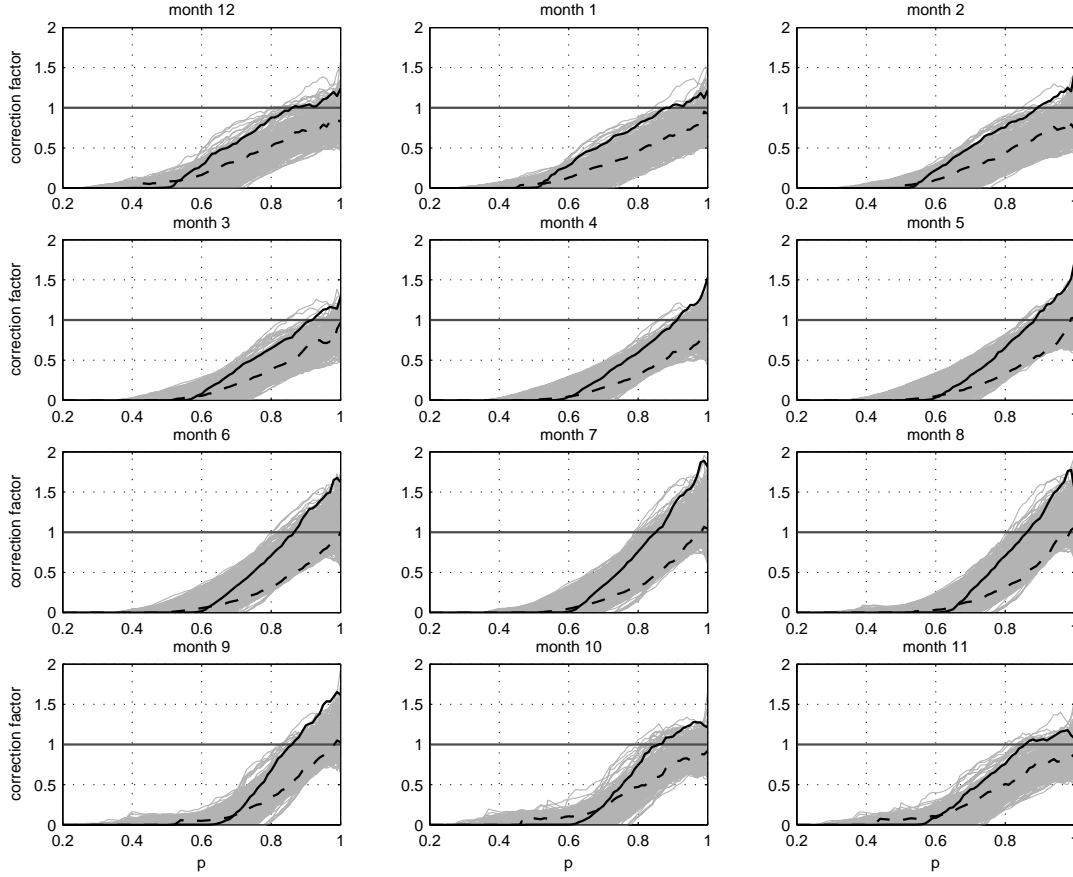
Figure 6: The monthly correction factors for precipitation

The performance of the correction method is evaluated especially concerning two aspects: the ability of reproducing the CDF and the ability of reproducing various statistical parameters representing the day to day performance of the correction method and of the model data.

The CDF validation gives information of the performance of the method and its ability to fulfill the general aim of the method. To test this CDF reproduction different periods are used for validation und evaluation. It has to be proven that the correction factors obtained from one period might correct the data of a different period in a similar way in a statistical sense.

If this first task is passed it has to be tested what the effects of the correction are on the data in general. Since the data are corrected they are also changed. They are changed to fit the CDF of the observations, but how these changes affect the data statistics in general has to be investigated as well. This second aspect of evaluation concerns the general statistical performance of the method with a closer view on statistical parameters and quantities characterizing of precipitation and maximum temperature.

General statistical performance

To compare the skills of the bias correction, a whole set of standard statistic parameters are calculated. In **Figure 7** several indices for precipitation and temperature maximum is shown. Always the fraction model versus observation is calculated. The indices are calculated for every single grid point and than averaged over the whole area. For precipitation (left side) a clear improvement for all quintile depending indices like mean, intensity etc, can be seen. All

indices are close to 1 which indicates a good fit with observations. Only for the very extreme values like the 3 maximum daily precipitation sums no improvement can be seen.

For temperature maximum (right side) for all indices an improvement can be seen. For the average values and not percentiles up to 80 the bias nearly vanishes. At the 95 percentile the bias is improved but still reaches more than 1.5 K and only a small effect is seen at the mean absolute error. This can be explained by no improvement of the temporal correlation.
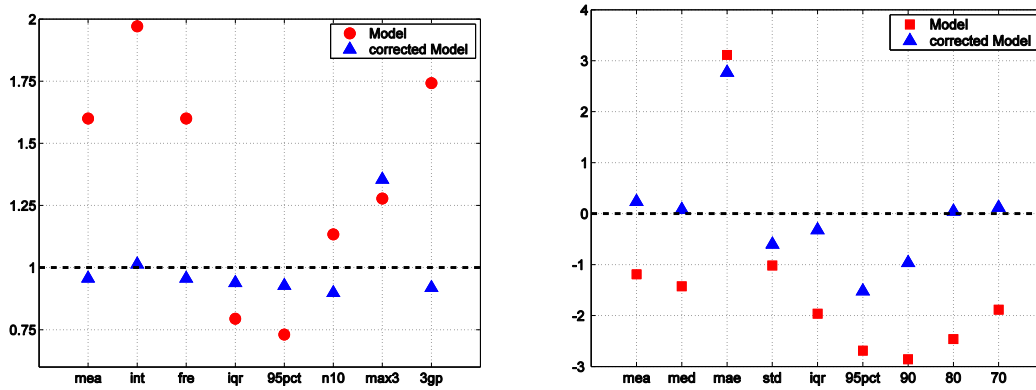

Figure 7: Comparison of the corrected model and the model performance. The left panel shows results for precipitation and the right panel for maximum temperature. Shown are a number of statistical indices characterizing precipitation. The values are fractions of the observational value. The indices for left side: m= mean, int=intensity, fre=frequency, iqr=interquartile range, 95pct=95 percentile, n10=number of days with more than 10 mm precipitation, max3= average of the 3 highest daily precipitation sums, 3gp=average of the 3 maximum gridpoints. The indices for right side: mea=arithmetic mean, med=median, mae=mean absolute error, std=standard deviation of the bias, iqr=interquartile range, 95pct (90,80,70) = 95 (90,80,70) percentile

For precipitation the situation is more complex and more difficult. This is due to the CDF of precipitation and due to the binary character of precipitation which separates days with precipitation and days without precipitation. On the one hand we have a continuous variable with a range starting from zero but on the other hand all variables with zero precipitation are distinctly different in character to all other days. We assume further that all models produce too many days with too little precipitation which is the case in general. The correction method now changes a number of days with little precipitation to zero. This is already a question worth change since the character of each of these days is substantially changed. But since these days are days with very little precipitation this can be tolerated. But since the number of days where this changes has to be done is a very big portion of all days these changes have to be somehow compensated by all the following values which leads to major changes in all statistics concerning only precipitation days.

The effect of the bias correction on the frequency to consecutive wet and dry day is shown in Figure 8. In the direct model output the frequency dry periods is highly underestimated. Even short dry periods are underestimated and it is getting worse the longer the dry periods are. This is clearly improved by the bias correction, but even then a slight overestimation of short dry periods and underestimation of long periods is given. For wet periods (right site) the original model overestimate long wet periods with more than 6 consecutive wet days. This bias vanishes after bias correction nearly totally.
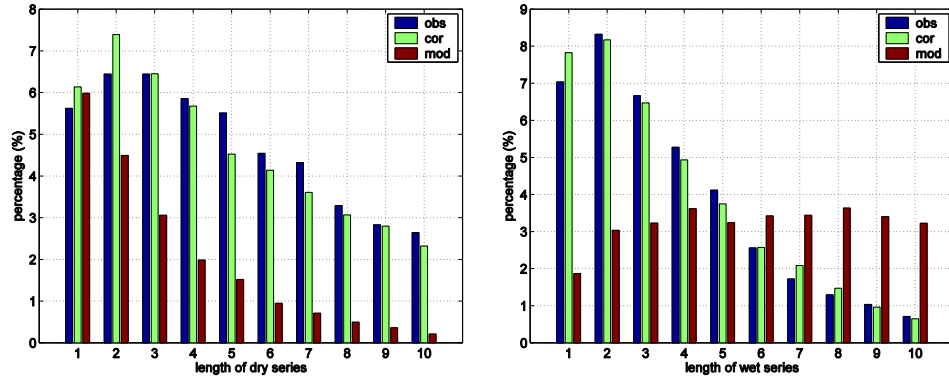
Figure 8: Series of consecutive dry and wet days. (left dry series, right wet series)

In Figure 9 and Figure 10 the frequency distribution of the biases before (left side) and after bias correction (right side) is shown. For temperature maximum biases of the model are nearly normal distributed but shifted to the left, which indicates a cold bias. After bias correction the biases are symmetric around zero and nearly 50 % of the biases are within ± 1 K and more than 80 % within ± 2 K.
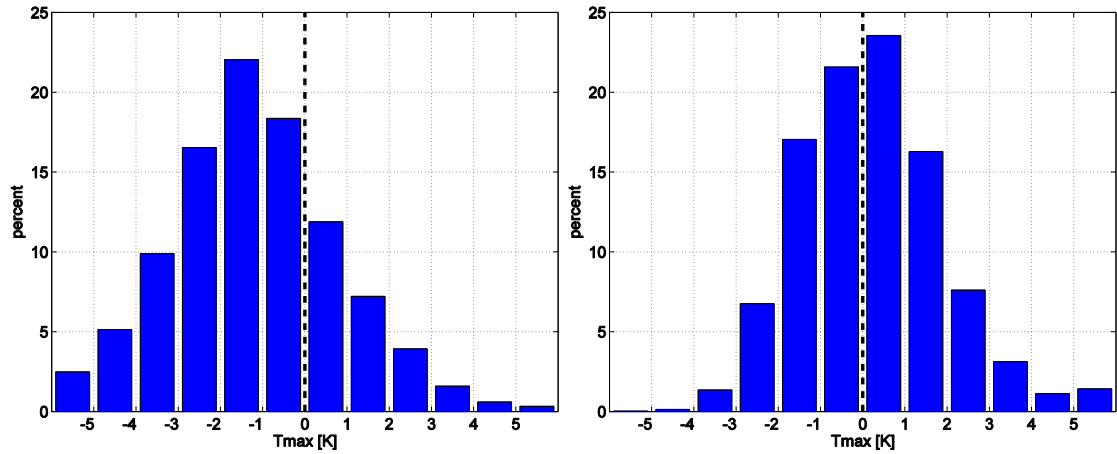

Figure 9: Histogram of the mean anomalies of the model maximum temperature (left) and the corrected model maximum temperature (right)

A similar improvement can be seen for precipitation. Here the model precipitation biases are shifted to the right side, which indicates an overestimation of precipitation. After bias correction more the 45 % of the biases are within ± 0.5 mm per day
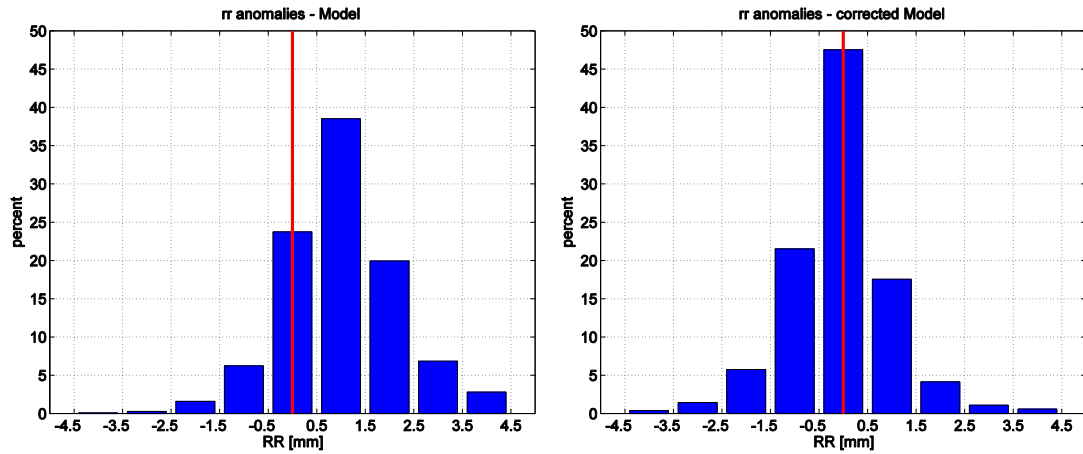
13

Figure 10: Histogram of the mean anomalies (mm per day) of the model precipitation (left) and the corrected model precipitation (right)

The spatial distribution of the maximum temperature bias for summer (JJA) of the model (left) and the corrected data (right) is shown in Figure 11. Large areas show a cold bias and the average bias is 2 K. After applying the bias correction the average bias in 0.4 K and for most parts of the area the biases are within $\pm$ 1 K.
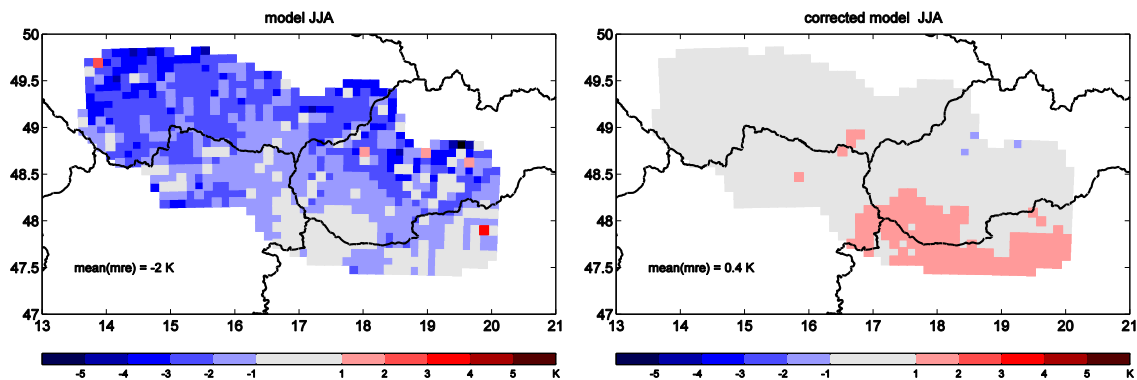

Figure 11: Spatial distribution of mean absolute errors of the monthly means for the model (left) and the corrected model (right) for maximum temperature.

The spatial distribution of the precipitation bias for summer (JJA) of the model (left) and the corrected data (right) is shown in Figure 12. Within the whole area the precipitation is overestimated by 50 %. After bias correction the area average bias is reduced to 3 %. But it can be seen that on grid point levels still relative biases of $\pm$ 25 % occur.
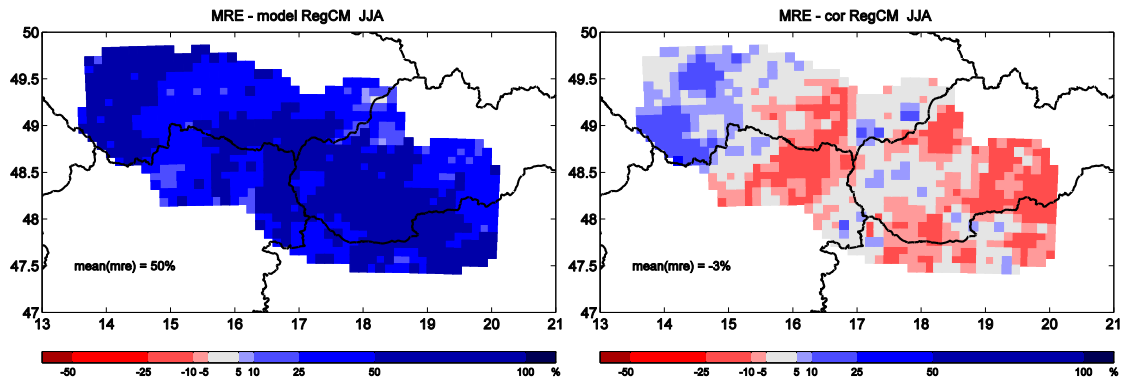
Figure 12: Spatial distribution of mean relative errors of the monthly means for the model (left) and the corrected model (right) for precipitation.

Application of the bias correction method

For application of this method to the climate change scenarios the regional climate control run 1961 till 2000 forced with GCM has been calibrated with observations and has been applied to the two scenario periods. Bias corrected data are available for the CECILIA Central Europe region on daily time step for

Temperature minimum
Temperature maximum
Relative humidity
Precipitation

Statistical postprocessing of RegCM3 and ALADIN regional climate models (CUNI)

Summary

At CUNI, a low-parametric alternative to the bias-corrective procedure applied at BOKU was investigated and used to modify data representing daily maximum and minimum temperature and daily precipitation. Its application is presented in the 'Bias correction' section for the outputs of the RegCM3 (run at CUNI) and ALADIN-Climate/CZ models (run at CHMI), driven either by the ERA-40 reanalysis or by ECHAM5 (RegCM3) and ARPEGE (ALADIN) global models - see deliverable D2.1 for details on the regional models. In the 'Localization' section, possibility of further localization of the RCM outputs is discussed and demonstrated for the RegCM model. Finally, bias-corrective and localization procedures have been combined, to provide very high resolution series with reduced systematic errors – a sample of the results is shown in the 'Combined postprocessing' section, and the concluding remarks are given in the 'Discussion and conclusions' section.

Bias correction

There are several strategies applicable for correcting the GCM/RCM outputs, from simple application of additive/multiplicative factors to more complex techniques based on deriving individual transformations for different sections of the data (e.g., for different quantiles). An overview of some of the approaches is presented by Deque (2007). Here, we apply a modified version of the method introduced by Piani et al. (2009).

First, both observed and simulated PDF of the variable of interest (daily maximum /minimum temperature or precipitation) were approximated by a low-parametric fit, individually for each grid point. Several distributions have been tested with regard to their ability to reproduce the shape of the temperature/precipitation distribution (normal, skew normal, log-normal, exponential, generalized exponential, GEV, gamma and generalized gamma). Two 3-parametric statistical distributions have then been chosen for postprocessing:

- **Skewed Gaussian distribution** was used to approximate the distribution of maximum or minimum temperature t

$$f(t) = \frac{2}{\sigma} \phi\left(\frac{t-\mu}{\sigma}\right) \Phi\left(\alpha\left(\frac{t-\mu}{\sigma}\right)\right), \tag{1}$$

where μ represents a parameter of location, σ parameter of scale and α controls asymmetry of the distribution; ϕ and Φ represent density and distribution function of the N(0,1) normal distribution.

- **Generalized gamma distribution** was applied to approximate the distribution of precipitation amounts p in non-dry days:

$$f(p) = \frac{\beta}{\Gamma(k)\theta}\left(\frac{p}{\theta}\right)^{k\beta-1} \exp\left(-\left(\frac{p}{\theta}\right)^{\beta}\right), \qquad x > 0 \tag{2}$$

16

where β and k are the shape parameters and θ is the scale parameter; Γ represents the gamma function. The fitting of the distributions (i.e., estimation of the parameters α, μ, σ and β, k, θ, respectively) was done by a gradient descent search for the maximum of the likelihood function. The relative number of days without precipitation (i.e., $p = 0$) was set to a constant value P, derived from the fraction of dry days in the calibration period of the simulated/observed data.

After obtaining the fits f, the respective CDFs were constructed through numerical integration of the PDFs (for precipitation, taking also the occurrence of dry days into account). The respective quantiles were computed with a step of 0.1%. By matching the quantiles representing the same position in the sorted datasets of RCM-simulated and observed values, corrective terms were defined as a function of the RCM-simulated t or p. The correction was created to be additive for temperature series (i.e., its value was calculated as a difference of the matching quantiles), and as the multiplicative one for precipitation. The precipitation values exceeding the 99.8% quantile were left in their original form; temperatures below the 0.2% quantile (over the 99.8% quantile) were modified by the correction obtained for the 0.2% (99.8%) quantile - this modification was used to reduce the effect of sometimes poor representation of the extreme tails of the target variables by the low-parametric fits.

Calibration of the corrective mappings was carried out for the period of years 1961-1985, validation for the period 1991-2000. The corrections were created separately for individual seasons (precipitation) or months (temperature). Identical spatial validation domain was used by both BOKU and CUNI teams (for more information on this domain and calibration/validation dataset generated from observations, see the report on deliverable D3.1).

The effect of the presented postprocessing procedure on the bias of the maximum temperature series produced by the RegCM and ALADIN models is illustrated in Figs. 12 and 13. The temperature fields are brought very close to the observed patterns by the correction and their cold bias almost disappears at most locations in the annual mean (Fig. 12). In some seasons, there is a residual bias even after the correction, but match to the observed data is substantially improved nonetheless. The bias is also generally comparable to or smaller than bias of the outputs of the statistical downscaling techniques (Fig. 13), described and studied in the deliverable D3.3. The results were similar for minimum temperature (not shown).

Systematic errors are even more of a problem for the RCM-simulated precipitation than they are for temperature. Fig. 14 illustrates this for the mean annual precipitation: While the RegCM model suffers from a very strong wet bias, ALADIN reproduces precipitation fairly well in a spatial average, but distinctly overestimates spatial variability. Application of postprocessing can remove both of these problems, even for individual seasons (Fig. 15). Again, the bias of the corrected data is no worse than for the series produced by statistical downscaling techniques. A substantial improvement can be observed not just in terms of mean values and spread of values, but also in a more realistic shape of the PDF in general (Fig. 16). Note especially how the severe underestimation of the number of dry days is amended, while the number of days with heavy precipitation is reproduced rather realistically.
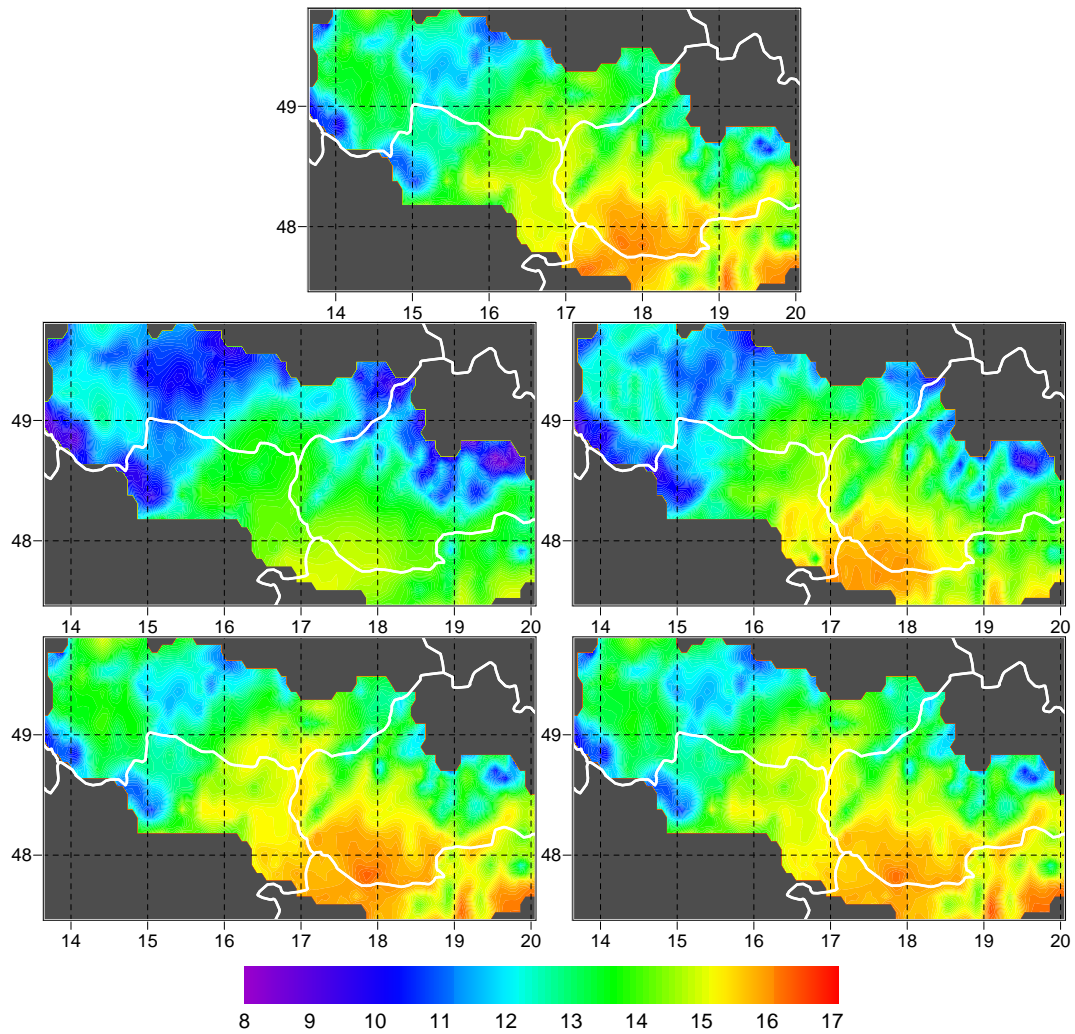
Figure 13: Annual mean of maximum daily temperature (°C): Observed **(top)**, raw output of the RCMs **(middle)** and outputs of the RCMs after correction **(bottom)**. Results for the RegCM model are shown on the **left**, for ALADIN on the **right**. The RCM runs were ERA-40-driven.
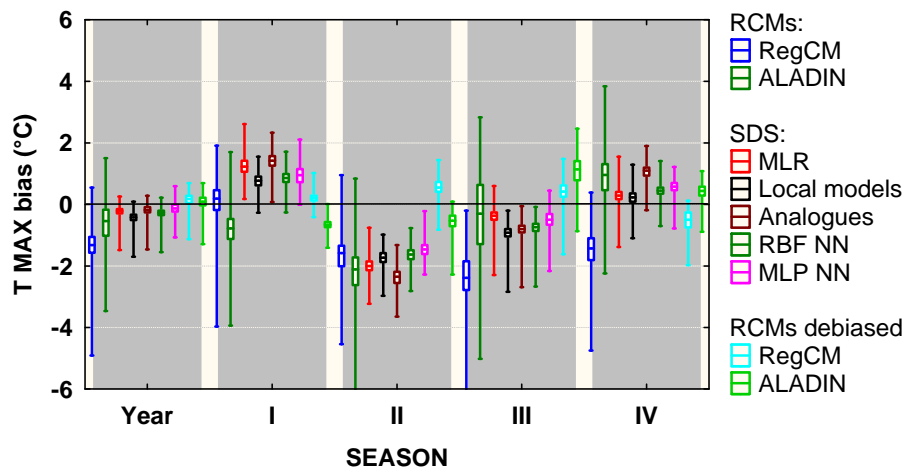

Figure 14: Bias of maximum daily temperature, for raw outputs of the regional models (RegCM. ALADIN), selected methods of statistical downscaling (MLR, Local models, Analogues, RBF NN, MLP NN – see deliverable D3.3 for details) and postprocessed outputs of the regional models. Seasons are denoted I (DJF), II (MAM), III (JJA) and IV (SON). The RCM runs as well as the statistical downscaling mappings were driven by ERA-40.

18

Figure 15: Annual mean of precipitation (mm/day): Observed **(top)**, raw output of the RCMs **(middle)** and outputs of the RCMs after correction **(bottom)**. Results for the RegCM model are shown on the **left**, for ALADIN on the **right**. The RCM runs were ERA-40 driven.
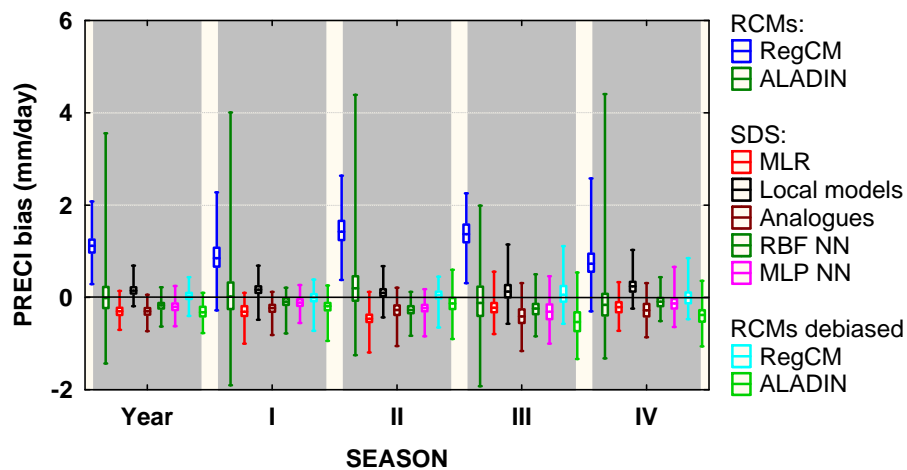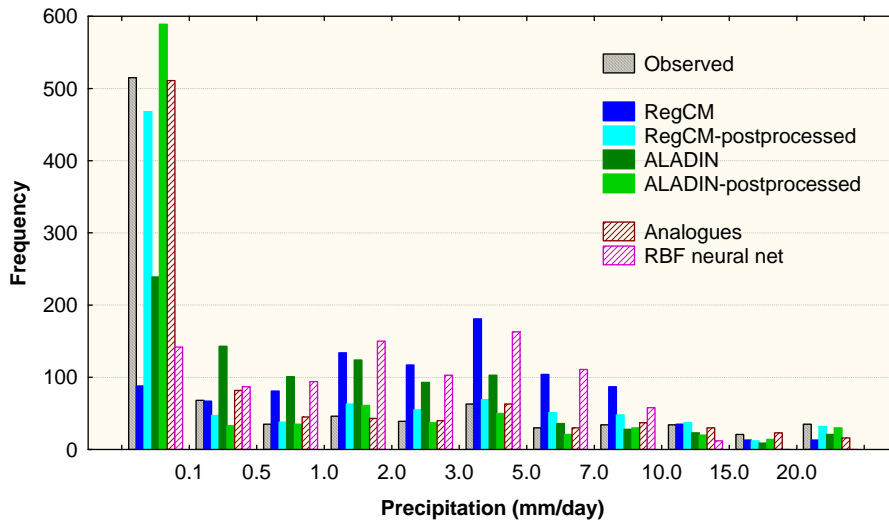

Figure 16: Bias of precipitation, for raw outputs of the regional models (RegCM, ALADIN), selected methods of statistical downscaling (MLR, Local models, Analogues, RBF NN, MLP NN – see deliverable D3.3 for details) and postprocessed outputs of the regional models. Seasons are denoted I (DJF), II (MAM), III (JJA) and IV (SON). The RCM runs as well as the statistical downscaling mappings were driven by ERA-40.

19

Figure 17: Distribution of values of daily precipitation in the series obtained by different downscaling methods, dynamical or statistical (situation for a grid point located at 49°00' N, 15°28' E) – JJA season. The RCM runs as well as the statistical downscaling mappings were driven by ERA-40.

By re-shaping the distribution of the target variables, representation of individual quantiles generally becomes significantly better. In Fig. 18 this is illustrated for the 60% quantile of precipitation (which is relatively close to zero in the observed data, but unrealistically high in the raw output of the RegCM model) and the 99 % quantile of precipitation (which is simulated with a minor tendency for underestimation; postprocessing introduces a mild change to correct this at most grid points).
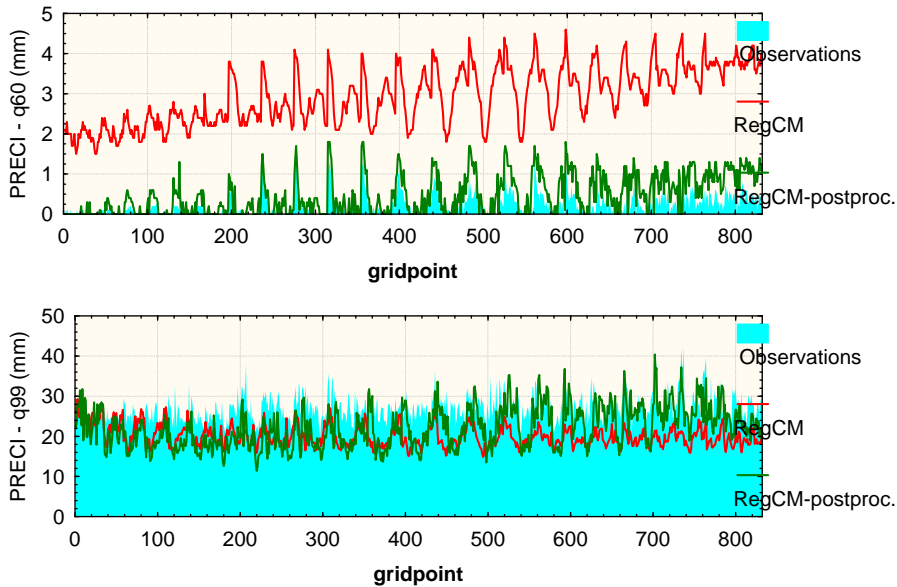

Figure 18: 60% **(top)** and 99% **(bottom)** quantile of daily precipitation in the JJA season, in observed data, raw RegCM (ECHAM5-driven) model outputs and corrected model outputs. The horizontal axis shows the identification number of the grid point in the validation domain.

20

Improved representation was also typical for the extreme tails of the temperature distributions. While the very highest and very lowest quantiles were often simulated with higher systematic error than means/medians, the applied correction was generally able to bring them back close to the observed values (Fig. 19), though its performance was sometimes worse for the ERA-40-driven RCMs. An improvement was also generally achieved for the measures of asymmetry of the PDF (Fig. 20).

An important feature of the applied corrections is the character of their eventual influence to the simulated changes of the climate variables. Even thought the form of the correction is identical whatever period of the data it is applied to, the resulting climate change signal may be amplified or reduced due to the fact that the corrective factors vary as a function of the values being corrected. Examples of such behavior are given in deliverables D3.4 (climate change scenarios for the period 2021-2050) and D3.5 (period 2071-2100).
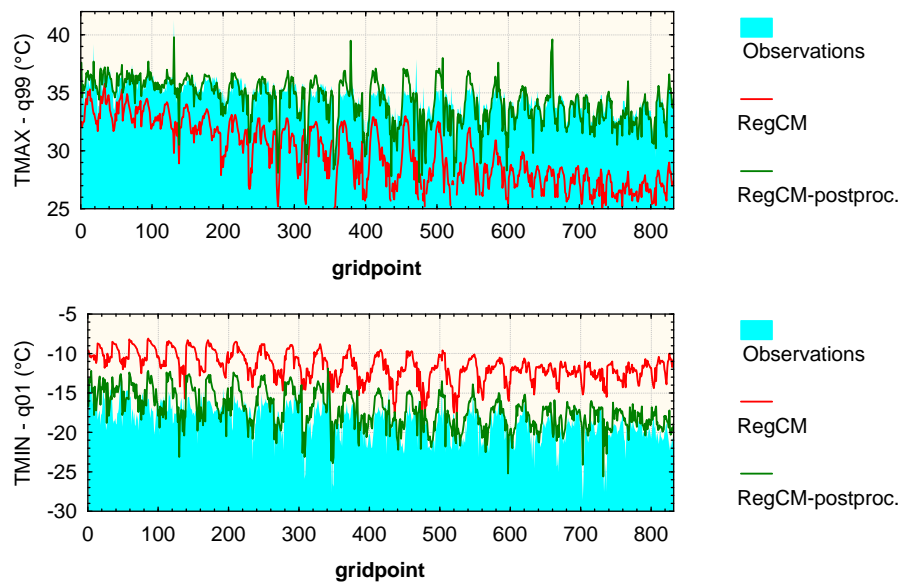

Figure 19: 99% quantile of maximum daily temperature in the JJA season **(top)** and 1% quantile of minimum daily temperature in the DJF season **(bottom)**, in observed data, RegCM (ECHAM5-driven) raw model outputs and corrected model outputs. The horizontal axis shows the identification number of the grid point in the validation domain.
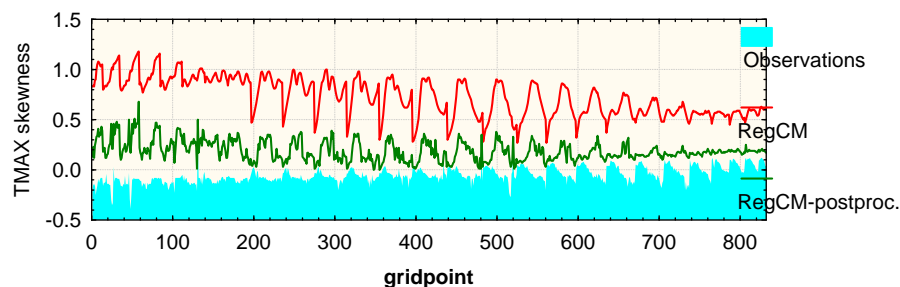

Figure 20: Skewness of maximum daily temperature in the JJA season, in observed data, RegCM (ECHAM5-driven) raw model outputs and corrected model outputs. The horizontal axis shows the identification number of the grid point in the validation domain.

Localization

The simulations run within the frame of the CECILIA project (using models RegCM and ALADIN-Climate, described in CECILIA deliverable D2.1) were carried out with the horizontal step of approximately 10 km. But even with the model grid this dense, the complexity of the surface and its influence to the atmospheric processes can be captured just partially. The seriousness of this problem varies with the geographic region - the predominantly flat areas are described rather realistically, a complicated terrain is unavoidably oversimplified. The Czech Republic is among the regions where the real terrain is still too complex to be accurately represented by a model, and the difference between real orography and orography in the 10 km grid is profound at some locations (Fig. 21). The same can also be said for most of the CECILIA Central-European validation domain.
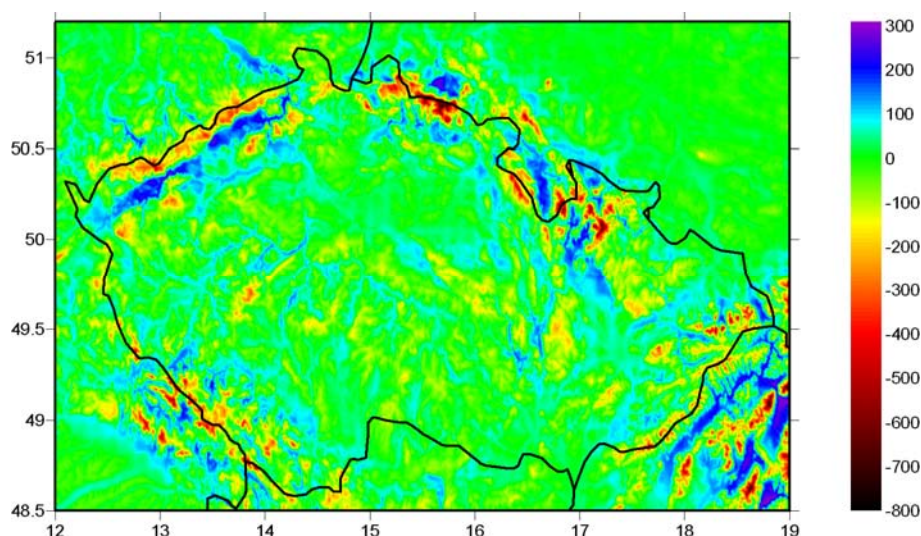

Figure 21: Difference (m) of the model orography used by the 10km RegCM run (CUNI) and real terrain height, described by the USGC 30" digital elevation map (obtained from http://www.usgs.gov/).

While the terrain (mis)representation is just one of the issues associated with a finite resolution of regional climate simulations, its effect is critical for many key climatic variables. In their approach to the RCM outputs localization, the CUNI team has focused on finding general mappings, able to compensate for the identifiable effects of altitude mismatch, and on possibilities of their combination with the bias-corrective procedures.

Because the resolution of the RCM simulations in CECILIA already exceeds the density of data directly available from the network of weather stations (every model grid point represents area of approximately $100 \, \text{km}^2$), no attempt was made to create individual location-specific transfer functions for each target site (such techniques of statistical downscaling were, however, developed for CECILIA deliverable D3.3). Instead, the applied approach to localization was aimed at finding a modification/correction of the model outputs, which:
- **is applicable for any arbitrary location**, not restricted just to sites with a pre-existing record of measurements of the target variable,
- **can be used for both present and future climate** and will not require any strong assumptions about the stationarity of the derived relations.

Several approaches to localization of RCM data have been tested and compared for the outputs of the RegCM model. The altitudinal dependence of temperature or precipitation was

employed as a primary base for the transformation of the model outputs to arbitrary points inside the model domain (including locations not identical with any of the nodes of the model grid), compensating for the oversimplified orography of the model. In addition to the basic altitude-based conversion, the results of localization were further combined with bias correction introduced in the previous section. The presented results were derived for the area of the Czech Republic, using the ERA-40-driven integration of the RegCM model (period 1961-1990); some of the tests were also carried out for CECILIA validation domain.

First, several types of localization based on global (i.e., related to the entire target region, approximately of the size shown in Fig. 21 or 25) or local (i.e., computed on a limited neighborhood of the target location) vertical gradient were tested. Various variants of the correction have been examined, involving localization of monthly mean or individual daily values, with monthly or daily values of vertical gradient employed for the correction of the altitude discrepancy.

A simple form of corrective mapping was used to derive the localized values:

$$x_{\text{localized}} = x + b\left(h_{\text{O}} - h_{\text{M}}\right), \tag{3}$$

where x denotes the model-produced value of the variable being localized, b is the global/local vertical gradient of x for the respective month/day and h_{O} and h_{M} represent altitudes of real (USGS) and model terrain, respectively. For points located between the nodes of the RegCM grid, h_{M} and x have been interpolated by the inverse distance weighting (IDW). The value of gradient b was computed globally or locally, using the gridpoints from the 100 km radius around the target location.

The model-simulated gradient b proved to be a good approximation of the vertical gradient observed in the real atmosphere for daily mean temperature (Fig. 22, left panel); for maximum and minimum temperature, the difference was somewhat stronger. The model-produced vertical gradient of precipitation exhibited significant differences from reality, though the mean annual value was captured fairly well (Fig. 22, right panel).

For gradients computed on daily basis, the deviations from reality were more distinct (especially for synoptic situations characterized by temperature inversion). Even so, in a temporal average, the results were similar to those obtained by localization on mean monthly basis.
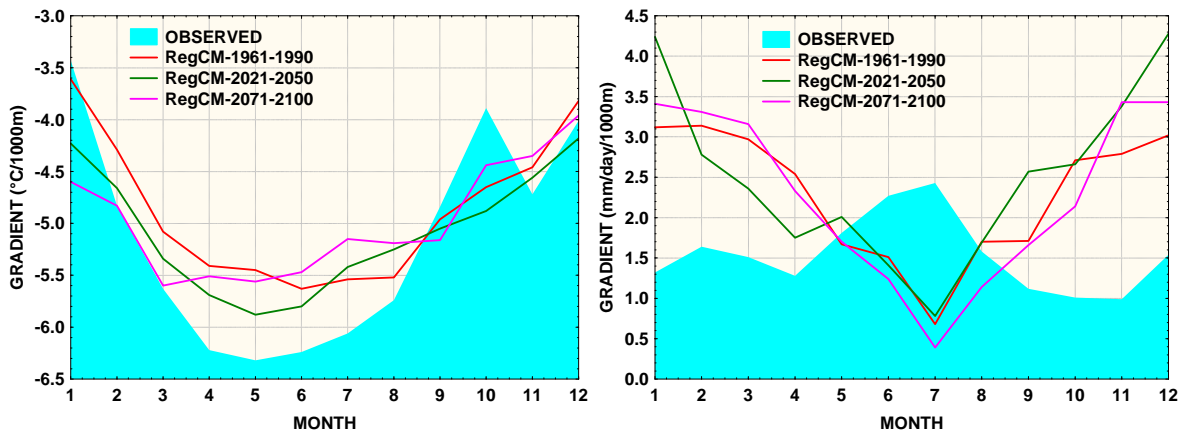

Figure 22: Mean vertical gradient of mean temperature **(left)** and precipitation **(right)** in the observed data and in the outputs of the RegCM model for different periods. Target area: Czech Republic

The linear correlation between monthly mean values of all three types of temperature and terrain elevations was very strong in the model data (see Fig. 23d for mean January temperature). Also, while there were certain spatial variations of the local value of the temperature gradient (Fig. 23c), the range of values was rather small over the Czech Republic. Because of this, the resulting temperature patterns followed terrain elevation very closely (Fig. 23b), and they conform to the patterns of observed temperature (e.g., Tolasz et al., 2007) quite well. For precipitation, the relation between monthly mean values and altitude was strong in the cold part of the year (Fig. 24), but much weaker during summer. As a result, the altitude-based correction did not dominate the field of localized precipitation in summer, and the spatial distribution of interpolated values was basically a product of the IDW interpolation alone.

The outcomes of the altitude-based localization were validated for a data set obtained at 25 Czech weather stations. Performance of the procedures based on daily data and on mean monthly data was similar; both offered improvement for temperature series at most stations in terms of RMSE or bias, but failed to better the series of precipitation.
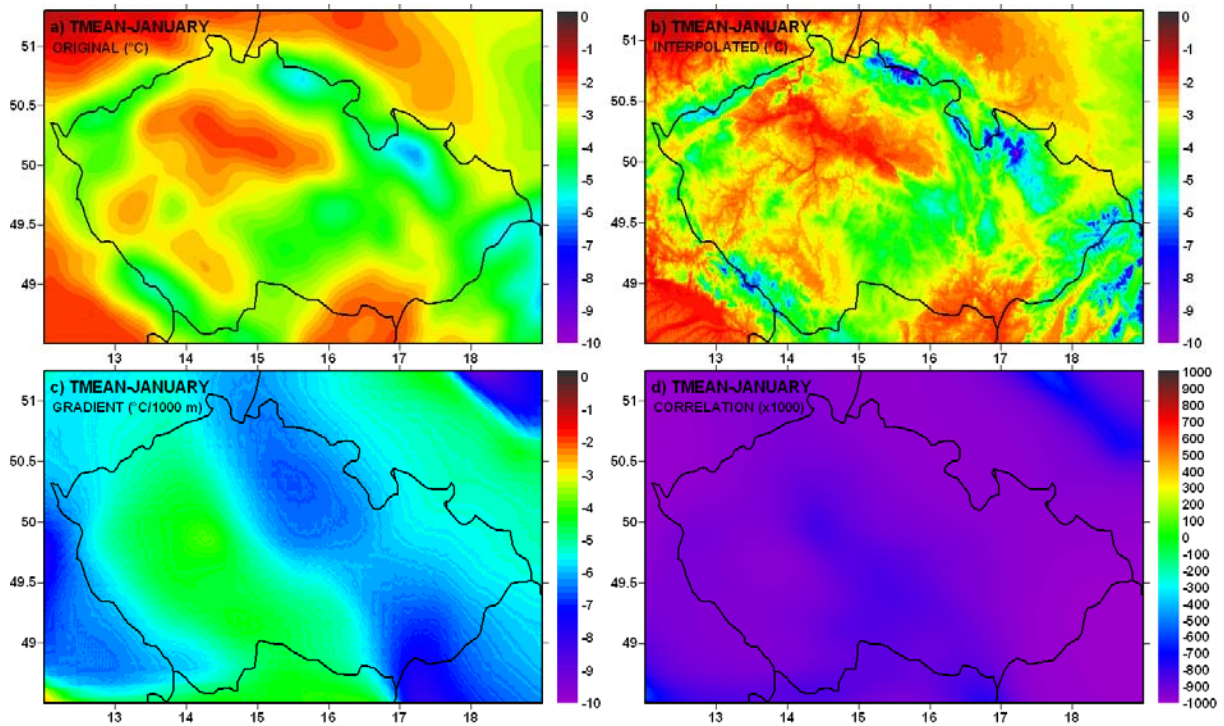

Figure 23: Monthly **mean temperature in January**, simulated by the RegCM model **(a)** and after the altitude-based localization of the model outputs **(b)**. The respective local values of the vertical gradient b of model temperature are shown as **(c)**, correlation (x1000) between temperature and altitude in the local neighborhood of the target location (100 km radius) as **(d)**.
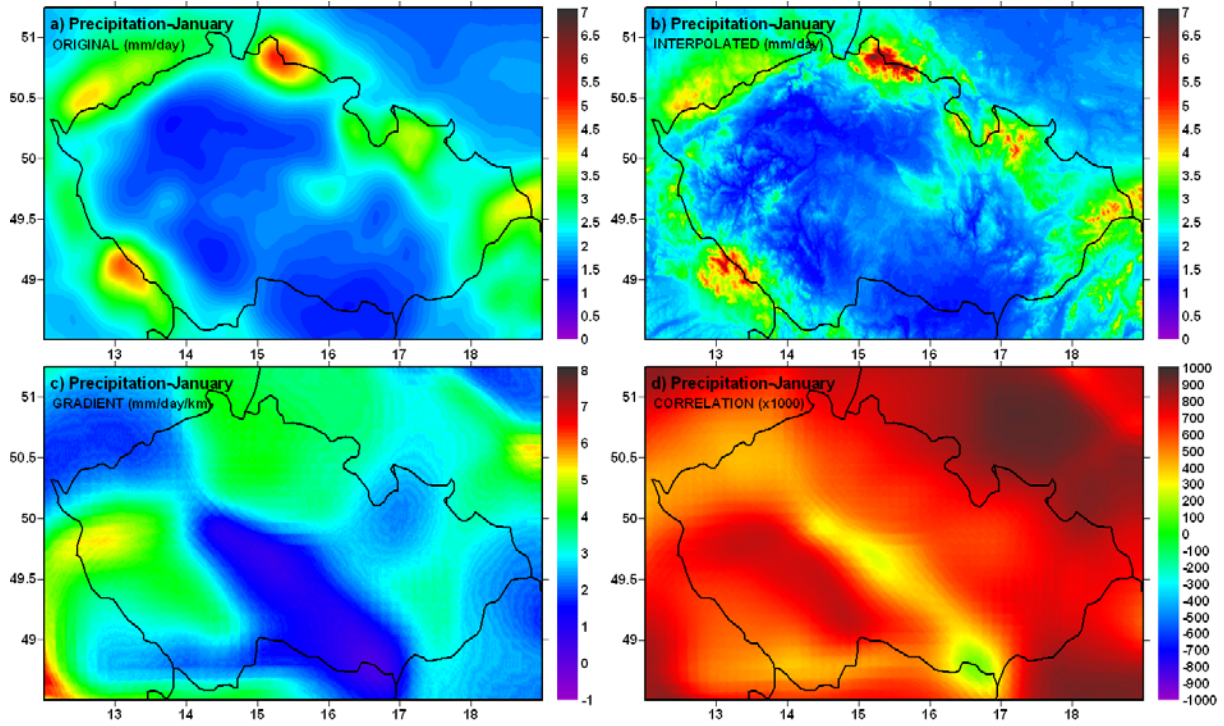
Figure 24: As Fig. 23, for **monthly mean precipitation in January**.

Combined postprocessing

To obtain fields of temperature and precipitation that are both devoid of major systematic errors and detailed enough to be used for studies at very fine scales, postprocessing and localization may be combined. Fig. 25 shows results of such a procedure applied to the outputs of the RegCM model, forced with ECHAM5, for annual mean of maximum daily temperature in the periods 1961-1990 (control climate), 2021-2050 (simulation for near future) and 2071-2100 (far future). The target area in this case was identical to the one employed for the postprocessing procedures in the chapter 'Bias correction'.

While generally successful, the procedure does not produce data which could be considered completely flawless. In particular, note several artifacts, visible especially in the temperature field simulated for the periods 2021-2050 and 2071-2100 in Fig. 25. Their occurrence is caused mostly by the insufficient local density of the observations, used for creation of the gridded validation/calibration data. In the future development of the postprocessing techniques, the severity of this problem may be reduced by using generalized, area-wise bias corrections instead of transforming the series for each grid point individually.
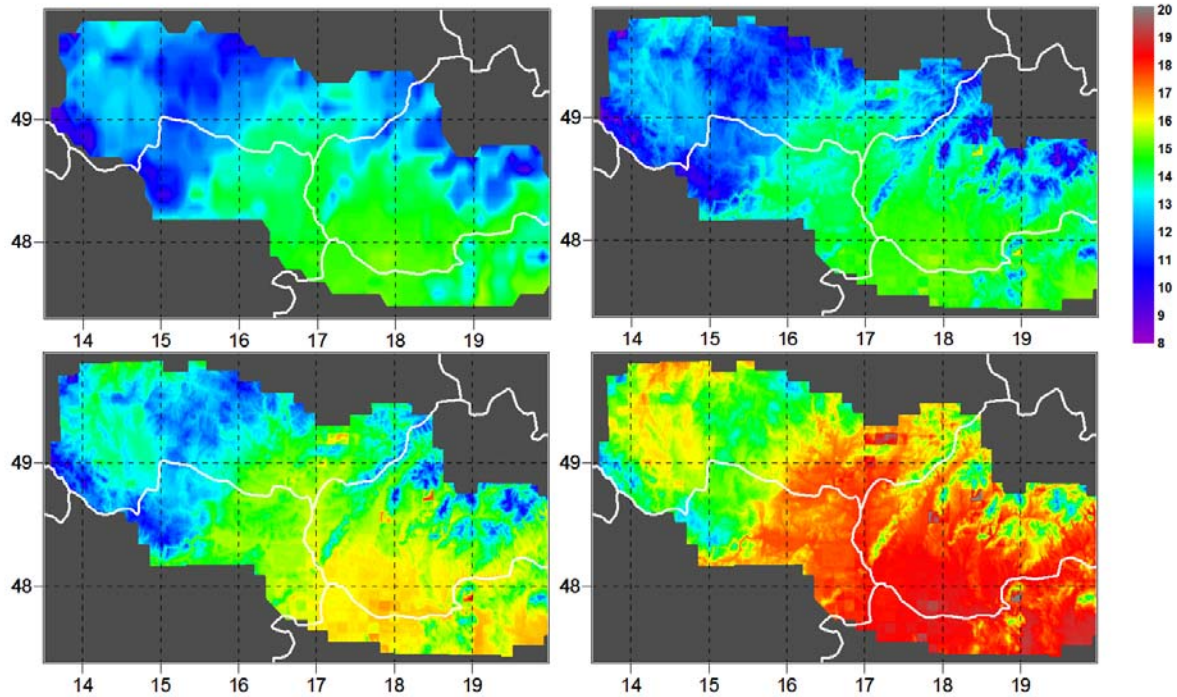
Figure 25: Annual mean maximum temperature (°C): Data simulated by the RegCM model after the bias correction **(top left)**, localized data for the control period 1961-1990 **(top right)** and the future periods 2021-2050 **(bottom left)** and 2071-2100 **(bottom right)**

Discussion and conclusions

While the quantile correction used at BOKU represents a more flexible mapping with more degrees of freedom (i.e., free parameters through which the mapping can be fitted), application of a function of such complexity may result in overfitting of the corrective function, which may then become unreliable for tasks involving extrapolation for conditions substantially different from the calibration ones (especially in case of more distant future time periods, when a significant shift of PDF is expected for some of the climate variables). Regardless of the details of the corrective procedures, both BOKU and CUNI approach seem to substantially improve the quality of the RCM outputs. But it should also be emphasized that there are problems still not addressed by the correction in the applied form(s): For instance, it does not specifically target and repair eventual unrealistic temporal structure of the series, characterized, for instance, by the shape of the autocorrelation function. Nor does it explicitly repair the discrepancies in the structure of spatial correlations, though the modifications applied at individual grid points may bring changes to the relations between the respective series.

The corrective algorithms, both the one applied at BOKU and at CUNI, focus on univariate conversions, performed separately for singular variables. This may lead to weakened physical consistency of the model outputs, and/or retaining the discrepancies contained in the raw model outputs. Future efforts in this field should therefore attempt to create multivariate corrective approaches, which will create datasets that respect the natural relations between different variables (and, at the same time, remove the respective inconsistencies in the raw outputs of the RCMs themselves).

The localization algorithm applied at CUNI proved to be rather successful for monthly/seasonal temperature means. This is due to the strong (and relatively linear) link between temperature and terrain elevation, which is reproduced quite well by the RCMs. For daily data, the simple linear correction seems inadequate in some synoptic situations, but substantial improvement was found on average. For precipitation, the localization in the presented form is less successful (especially in summer and for daily data), and more complex methods will probably be needed in the future (e.g., taking local circulation into account, as well as shape of local terrain instead of just elevation).

References:

Deque M. (2007): Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values. Global and Planetary Change 57, 16–26

Piani, C., Haerter, J.O., Coppola, E. (2009): Statistical bias correction for daily precipitation in regional climate models over Europe. Theoretical and Applied Climatology, in print

Tolasz R. et al. (2007): Climate Atlas of Czechia, Czech Hydrometeorological Institute & Palacky University, Praha, Olomouc, 255 pp.

Localization of the monthly mean temperature and daily precipitation using ALADIN simulations over Bulgarian domain (NIMH)

Localization method

By reducing the resolution of RCMs, some improvements may be expected due to more realistic impact of land surface fields and topography, reduced approximation error and improved dynamics. Despite this, discrepancies between real and model topography, real and model terrain, land-sea mask, etc. still remain.
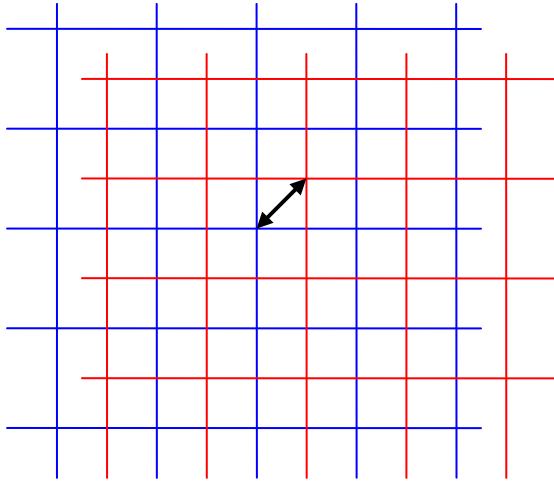
These inconsistencies are minimized by applying special transformation of the interpolated field to reduce the interpolation error.

Suppose an interpolation of any field $\mathbf{F}$ by a given operator $\mathbf{A}$ from one grid (blue) to a staggered one (red). Let us mention this operation by $\mathbf{A}^+$ and by $\mathbf{A}^-$ the backward one. Applying both of them we will obtain another value of the field for the same grid point:

$$\mathbf{A}^- (\mathbf{A}^+ \mathbf{F})$$

The difference between the initial and resulted values we call *'interpolation error' Ierr:*

$$\mathbf{Ierr} = \mathbf{F} - \mathbf{A}^- (\mathbf{A}^+ \mathbf{F})$$

Now we define the problem:

For a given interpolation operator $\mathbf{A}$ find transformation $\mathbf{B}$ of $\mathbf{F}$ so that:

$$\mathbf{B} \mathbf{F} - \mathbf{A}^- (\mathbf{A}^+ \mathbf{B} \mathbf{F}) = \mathbf{min}$$

The localization consists of the following steps:
1. Transformation of the field **F** by **B**.
2. Applying of interpolation to the observation point using **A.**
3. Applying reverse operator of **B**.

We use bilinear operator **A. B** was defined as linear combination of the elevation **h**, **h³** and derivatives of h: $\mathbf{d_x h}$, $\mathbf{d_y h}$ (this terms where obtained by preliminary analysis using stepwise regression):

$$\mathbf{B\ F = F + \alpha h + \beta h^3 + \gamma d_x h + \delta\ d_y h}$$

It could be proved, that the maximum error of the interpolation over the whole area is achieved when the grids are staggered equidistantly. The operator $\mathbf{A^- (A^+ BF)}$ is:

$\mathbf{A^- (A^+ B\ F) = (Bf_{i-1,j-1} + Bf_{i+1,j-1} + Bf_{i-1,j+1} + Bf_{i+1,j+1} + 2*(\ Bf_{i-1,j} + Bf_{i+1,j} + Bf_{i,j-1} + Bf_{i,j+1}) + 4*Bf_{i,j})/16.}$
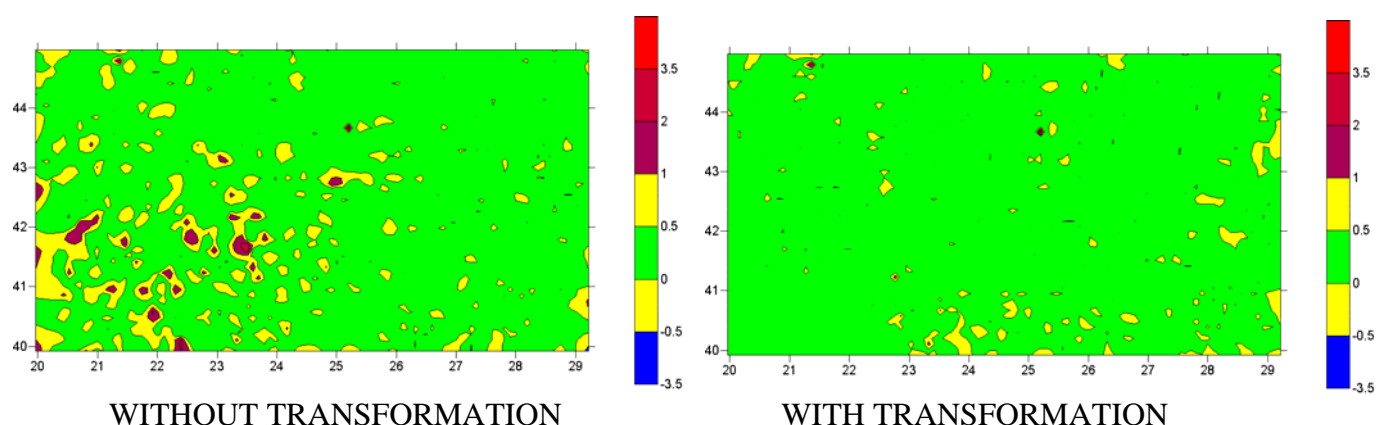
To avoid dependence of latitude and longitude the area was separated on several sub areas.

On the next figures the effect of this method is illustrated. For this we used daily runs for June 2007. The color scale for precipitation is in mm/day and Celsius degrees for temperature.

MONTHLY PRECIPITATION

WITHOUT TRANSFORMATION   WITH TRANSFORMATION

MONTHLY MEAN TEMPERATURE

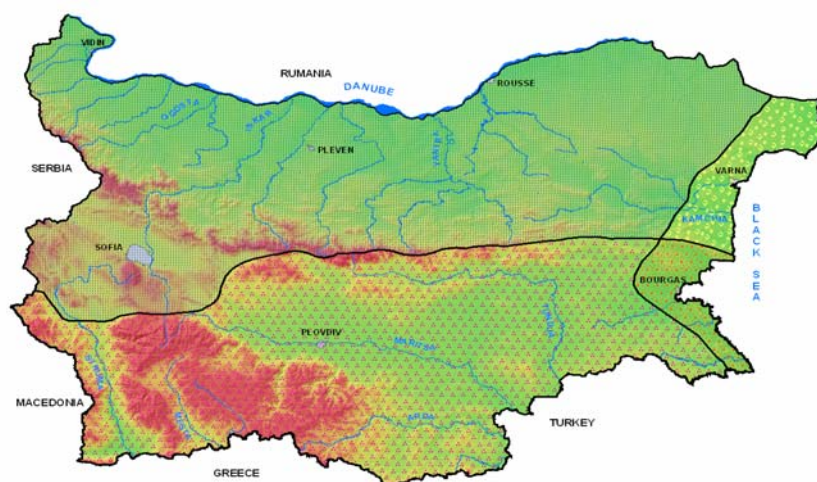WITHOUT TRANSFORMATION · WITH TRANSFORMATION

The above figures illustrate typical position of the errors. They are generally in narrow valleys and complex mountains – with several massifs like the Rhodope Mountains.

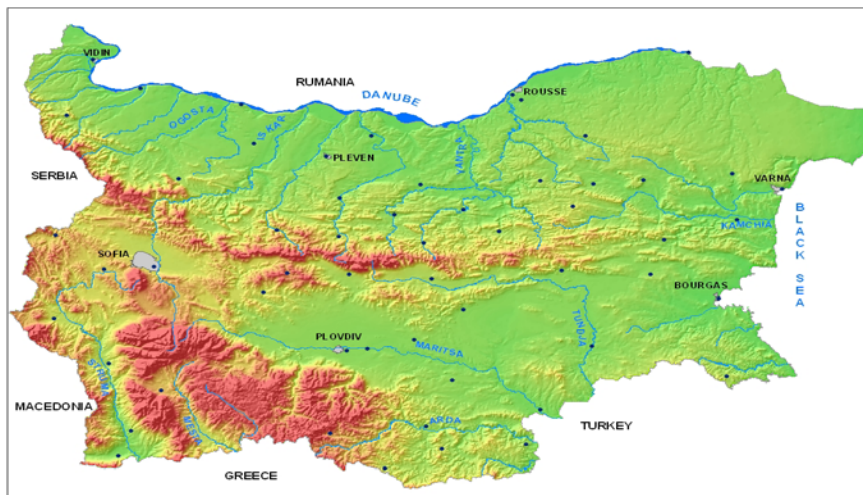## Verification of the localization method with ERA40 coupling of ALADIN RCM

Climatology

The barrier effect of the Balkan Mountains is felt throughout the country. On the average, northern Bulgaria is more then one degree colder and receives annually about 190 mm precipitation more than southern Bulgaria. Black Sea is too small to be a primary influencing factor of the country's weather; it only affects a narrow zone along the cost:
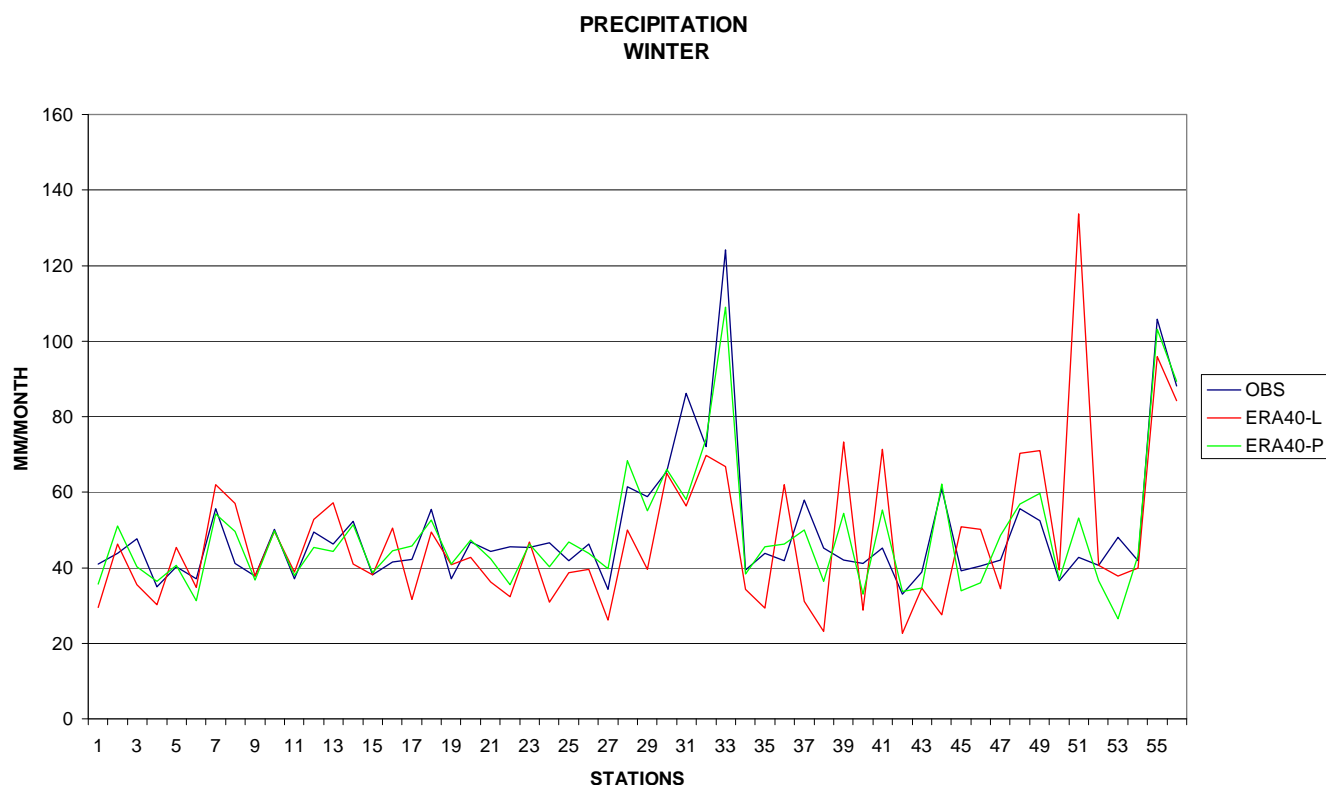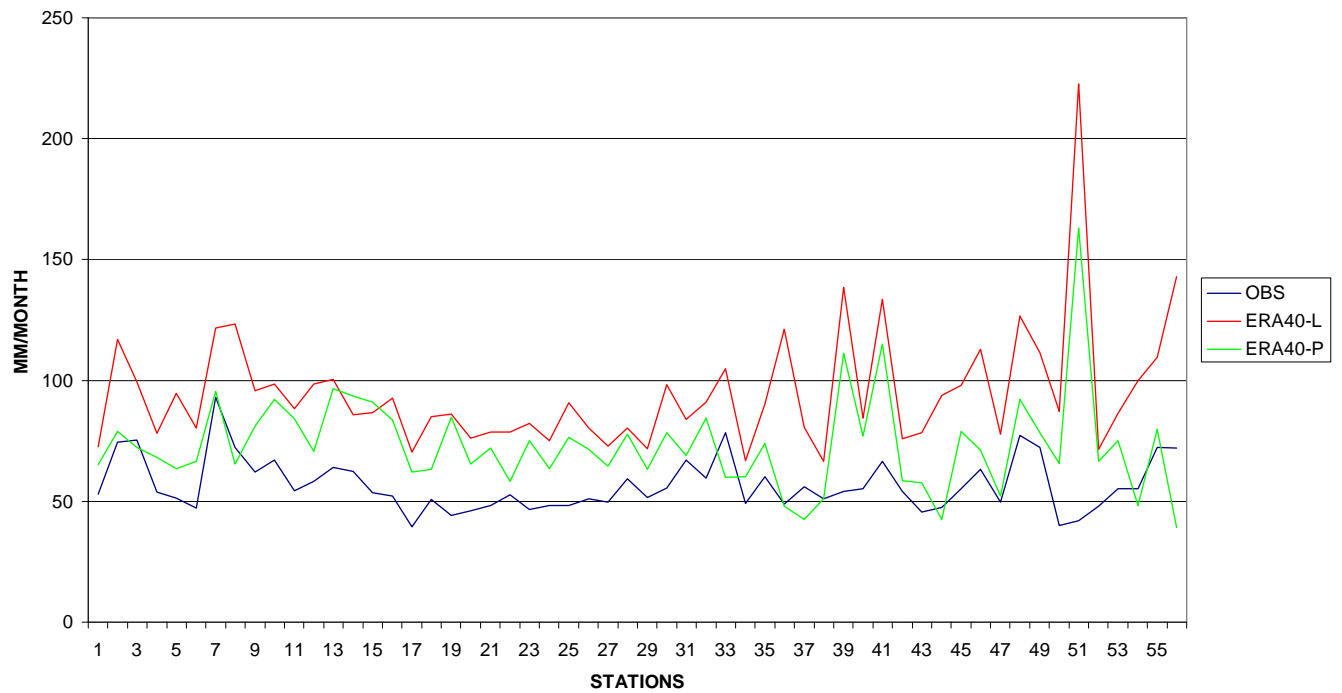

Verification dataset.

The problem is that we have not observation network of 10 km. We decided to use observations from stations situated in places affected by the land and topography characteristics and proportionally distributed at the mentioned above. 3 regions. An important criterion is the quality and the period without breaks. We selected 56 stations shown below.
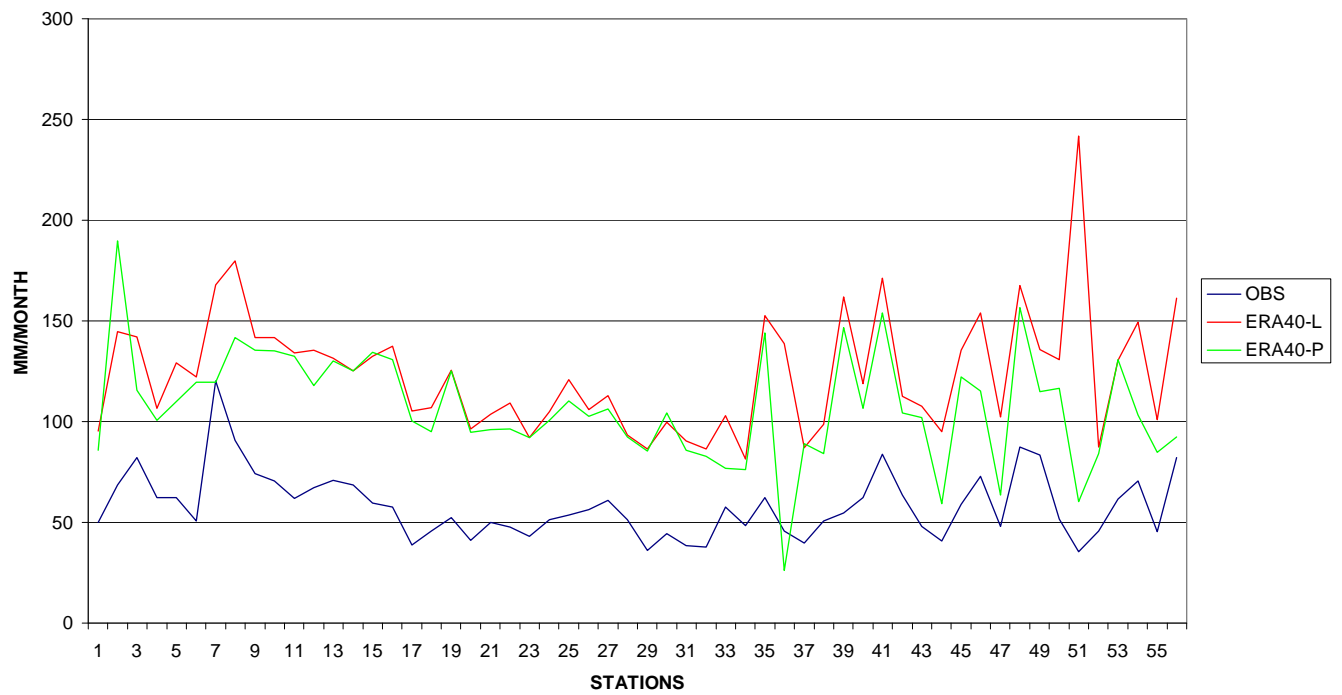
On the figures below observation (monthly mean temperature and monthly precipitation) are mentioned by blue color. The localization by direct bilinear approximation is indicated by red color and letter 'L'. The green color stays for optimized bilinear interpolation indicated additionally by 'P'. The biggest errors are located in the stations located in narrow valleys as it was mentioned above.
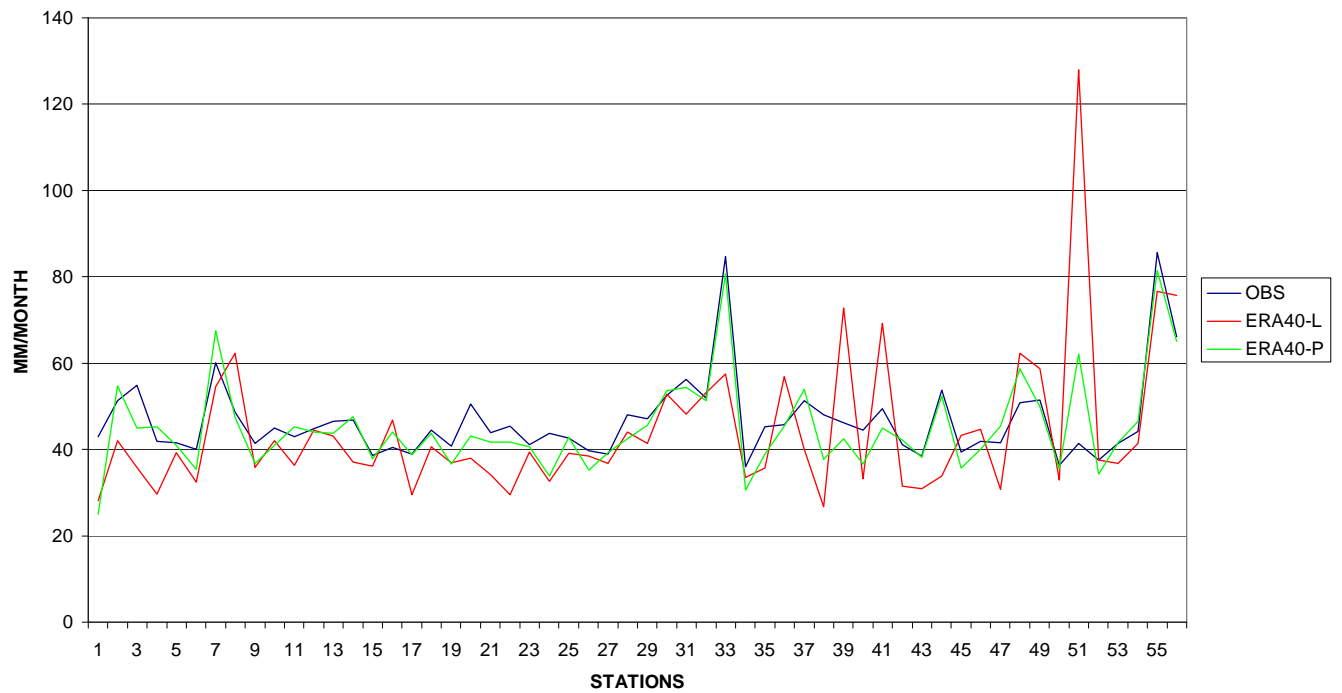
PRECIPITATION
WINTER

**PRECIPITATION
SPRING**

**PRECIPITATION
SUMMER**

PRECIPITATION
AUTUMN

TEMPERATURE
WINTER

TEMPERATURE
SPRING

TEMPERATURE
SUMMER

TEMPERATURE
AUTUMN