Originally published as:

Orlowsky, B., Gerstengarbe, F.-W., Werner, P. C. (2008): A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM. – **In**: Theoretical and Applied Climatology, Vol. 92, Issue 3-4, pp. 209-223

DOI: 10.1007/s00704-007-0352-y

A resampling scheme for regional climate simulations and its performance compared to a dynamical RCM

B. Orlowsky* (corresponding author), F.-W. Gerstengarbe and P. C. Werner

Potsdam Institute for Climate Impact Research

With 11 Figures and 1 Table

*E-Mail: Boris.Orlowsky@pik-potsdam.de

Summary

A new statistical method for regional climate simulations is introduced. Its simulations are constrained only by the parameters of a linear regression line for a characteristic climatological variable. Simulated series are generated by resampling from segments of observation series such that the resulting series comply with the prescribed regression parameters and possess realistic annual cycles and persistence. The resampling guarantees that the simulated series are physically consistent both with respect to the combinations of different meteorological variables and to their spatial distribution at each time step. The resampling approach is evaluated by means of a cross validation experiment for the Elbe river basin: Its simulations are compared both to an observed climatology and to data simulated by a dynamical RCM. This cross validation shows that the approach is able to reproduce the observed climatology with respect to statistics such as long-term means, persistence features (e. g., dry spells) and extreme events. The agreement of its simulations with the observational data is much closer than for the RCM data.

1 Introduction

Regional climate modeling has received increased attention in recent years. This is due to the fact that the effects of climate change vary strongly between different regions (Houghton et al. 2001), meaning for example heavier precipitation events together with flooding in one case and more severe dry spells and heat waves in another. This variety makes it necessary to analyse the changes in detail on a regional scale. Spatially highly resolved simulations of future climate developments are therefore needed as a basis for such impact studies.

One common approach to obtain these simulations is to translate output from Global Circulation Models (GCMs), typically operating on horizontal resolutions of 100–200 km, to a finer scale of, say, 10 km or to specific locations, e. g., meteorological stations. This task of *downscaling* raises two questions: First,*what* to downscale, and second, *how* to downscale it.

The first problem is concerned with the relation between explanatory and dependent variables. These notions mean that the evolution of an explanatory variable (for climate modeling often referred to as *forcing*) can be used to explain the evolution of a dependent variable. In the downscaling context, the explanatory variables are extracted from GCM output. The dependent variables are the local weather variables, which are to be predicted from the GCM output.

The chosen explanatory variables must be sufficiently relevant for the dependent variables of interest. For instance, if the rather complex precipitation patterns in an alpine region are to be obtained by a downscaling method, a sea level pressure field – although certainly linked to processes relevant for precipitation – will not by itself explain the fine-scale features of the spatial precipitation distribution (Beersma and Buishand 2003). Additionally, the chosen explanatory variables should be simulated reliably for the region of interest by the GCM. This is not necessarily given, as the vast literature about GCM validation shows (see e. g., Benestad 2004, Déqué 2004, Murphy et al. 2004, Fox-Rabinovitz et al. 2006, Ruiz-Barradas and Nigam 2006): The overall picture is ambiguous and does not give any clear indications on which GCM-simulated variables regional downscaling could be based.

For the second question, that is, how to downscale, a wide range of methods has been developed, using both dynamical (commonly referred to as Regional Climate Models, RCMs) and statistical approaches.

RCMs model physical processes as GCMs do, but on a finer spatial and temporal scale. Usually, GCM output defines the boundary conditions for these simulations. Recently, RCMs have improved substantially and are more and more used to produce climate change scenarios for impact studies. As their simulations are based on representations of physical processes, they are in particular able to deal with situations differing from the present climate. On the other hand, this general applicability is somewhat limited by the intrinsic parametrisations of sub-scale processes, which are often based on observational data statistics.

The applicability of statistical approaches in turn is per definition limited to similar climates – a serious drawback in times of climate change, which also concerns the new statistical method for regional climate simulations, which is introduced in this paper. Its performance is tested by means of a cross validation experiment. The climatology of a 25 years time span of the observation period is simulated and compared to the actual observations. As for the same time and region also a simulation by a dynamical RCM exists, the performance of the statistical model can be evaluated not only with respect to the observed climatology but also in comparison to the RCM.

In statistical downscaling, three different strategies can be identified: (1) transfer functions, making use of functional dependencies between explanatory and dependent variables (e. g., Murphy 1999, von Storch and Zwiers 1999, Zorita and von Storch 1999); (2) weather generators, which exploit stochastic models for variables of interest and generate simulated series by sampling from them (e. g., Semenov and Barrow 1997, Wilks 1999); (3) weather type schemes, using relationships between large-scale circulation regimes and local weather, which allow for the translation of changes in circulation regime statistics into changes of weather statistics. See Enke and Spekat (1997) for a combination of weather types and transfer functions for downscaling. All schemes rely on observed time series and aim at generating simulated series for a period of interest.

A special case of weather type schemes are analogue methods, to which the approach presented here is similar. They are typically used to generate simulated series of surface weather data which are compatible to a series of GCM-simulated large-scale circulation patterns: for the simulated circulation pattern of each time step they scan a pool of observed patterns, searching for the most similar and its concurrent surface weather (e. g., Timbal and McAvaney 2001, Beersma and Buishand 2003). The elements of the simulated series are obtained by conditional resampling from the pool of surface weather observations, where the resampling is conditional on the large-scale circulation fields. See Wilby and Wigley (1997), Wilby et al. (1998), Zorita and von Storch (1999), von Storch et al. (2000) for general reviews.

Like the analogue approaches, the scheme presented in this paper constructs simulated series by resampling from elements of the observation series. However, it is constrained by a forcing which is much simpler than series of daily circulation patterns as for the analogue methods: It consists only of the two parameters of a regression line, which the annual means of a characteristic climate variable have to feature. For the application in this paper, the characteristic variable is temperature.

Section 2 explains the scheme in detail, first for single station simulations only, which is then extended to multi station or regional simulations. Results from the cross validation experiment and the comparison to the performance of an RCM are shown in Section 3. Section 4 summarises and gives a short outlook.

2 The resampling scheme

Similar to the analogue methods, the approach presented in this paper is based on the assumption that weather states from segments of the observation period may occur again or very similarly during the simulation period. Hence, simulated series are constructed by resampling from segments of observation series, consisting of daily observations. The obvious advantage of such resampling is that the resulting series consist of observations (combinations of temperature, precipitation, pressure, etc.), which are insofar physically consistent, as they once were real world observations. The same holds for the spatial fields of the meteorological variables, as the series of the simulated fields is obtained by resampling from a pool of observed fields.

However, in contrast to the analogue methods, this resampling is not conditioned on daily series of GCM-simulated circulation patterns, which in general cannot be considered as a reliable explanatory variable (e. g., Mitchell and Ewins 2003, Murphy et al. 2004). The idea is instead to constrain the resampling by a very simple forcing, which can easily be checked for plausibility and does not depend on a single (possibly inaccurate) GCM simulation. As the only external constraint to the simulated series at a given location, the two parameters of a regression line (e. g., mean and slope) are prescribed, which the simulated annual means of a characteristic climate variable at this location have to feature. The characteristic variable is to be chosen such that it captures the essential climate variability of the region of interest. E. g., for the data from the Elbe river basin, which is used for the cross validation in Section 3, temperature is a good choice. The forcing in this case describes long-term level and (linear) increase of temperature over the simulation period.

Figure 1 illustrates this principle together with its external constraints: given are a time series of observations (here: temperature) and a regression line for the simulation period. From the observations, a simulated time series covering the simulation period is assembled, the annual means of which feature the prescribed regression line.

[Fig. 1 about here.]

This kind of forcing of course does not completely determine the simulated series. Therefore, a set of heuristic rules for the resampling is defined, which ensure that the resulting series exhibit realistic properties such as annual cycles and persistence. Stochastic elements in the resampling procedure make every simulated series a realisation of the population of all series compatible to the prescribed regression parameters and the set of heuristic rules. The bandwidth of this population, i. e., the range of possible climate developments under the assumption of the given linear development of the characteristic variable according to this approach, can be estimated by generating a large ensemble of simulations.

Generating a simulation can thus be seen as defining a date-to-date mapping, which each day of the simulation period assigns a day of the observation period and the concurrent observations. It should be noted that obtaining such a mapping is equivalent to obtaining simulated series of all observed variables, even if the construction of the mapping only uses a single climate variable. The mapping is constructed in a heuristic way which makes sure that the resulting series, besides featuring the prescribed regression parameters, exhibit properties such as realistic annual cycles and persistence. The notion of "realistic" in this context refers to what is known from the observation period. As the prescribed regression parameters are the only outer constraint, all other properties which make a simulated series realistic have to be checked against the observation series. Here, "realistic" therefore means that certain properties of simulated and observed series are statistically not distinguishable.

2.1 Single station simulations

The date-to-date-mapping is constructed in two steps, which operate on different time scales, the first on the time scale of years, the second on the shorter time scale of blocks of 12 days. They are, together with the preparation of the input data, illustrated in Figure 2, and will be described and motivated in the next paragraphs.

[Fig. 2 about here.]

Input: In preparation, the observational data is organised in year-wise segments (from January 1st to December 31st) and in sliding blocks of 12 days. These blocks of the observation period are grouped into n_c classes of similar blocks, "similar" referring to the $\mathbb{R}^{12}$-tuple of their daily temperatures, by means of a standard cluster analysis (see Appendix A). Each of the resulting n_c classes is represented by the centroid of its $\mathbb{R}^{12}$-tuples.

Step 1: The first step (grey shaded in Figure 2) generates a simple rearrangement of entire calendar years from the observation period, thereby producing a mapping which leaves the sequences within the years unaltered. Series generated by this first step are sure to exhibit realistic annual cycles and persistence, as the sequences within the years are simply copied from the observed series. The series are generated as follows:

1. The observed series are cut into pieces of one year length, always starting on January 1st.

2. These pieces are randomly rearranged by drawing with replacement. A large sample of such rearrangements is generated, each consisting of as many years as the simulation period. Such a rearrangement is shown in Figure 3, where for the simulation period 2007–2016 years from the observation period (1997–2006) are randomly rearranged (note the numbers indicating the years at the bottom).

3. From this sample, the rearrangement is chosen, which is the closest to the prescribed regression line. This is found by minimising the distance between the parameters of the regression line of each rearrangement and the parameters of the prescribed regression line over the sample.

4. If this rearrangement consists of more days than the simulation period (this is possible due to leap years), the extra days are removed from the end of the rearrangement. Otherwise, lacking days are appended from the beginning of the year which is the last in the rearrangement.

[Fig. 3 about here.]

Step 2: The second step operates with blocks of 12 days length. The use of blocks instead of single days makes sure that the weather sequences of the simulated series are realistic at least within the blocks, again simply because they are just copied from actually observed sequences. By experimenting with different block lengths, the length of 12 days is found to essentially capture the persistence of the observation time series from the Elbe river basin. It aims at ameliorating the series from **step 1** such that the resulting series matches the prescribed regression parameters. To this end, it exchanges selected blocks from this series with appropriate blocks from the observation series. This selection/replacement procedure is iteratively repeated until the prescribed regression parameters are reproduced.

Step 2 a: The first part of the second step in Figure 2 consists of identifying blocks which are to be replaced, namely those which contribute too much to the mismatch between the regression line of the series from **step 1** and the prescribed regression line. "Too much" in this context is detected as follows:

1. From the series of the characteristic climate variable from **step 1**, a synthetic series is generated by shifting the **step 1**-series year by year such that the regression line of the resulting annual means corresponds to the prescribed parameters. At this stage, two series of the characteristic variable exist. A few exemplary years of them are shown in Figure 4, grey for the series of **step 1**, blue

for the synthetic series. The figure also shows the annual means (dots) of the respective series. The first exhibits realistic annual cycles and persistence, the second, not very different, fits the prescribed parameters exactly.

2. Each of the consecutive 12 days-blocks is compared with respect to these two series. If a block is similar relative to both series, it is kept from the series of **step 1**. For the comparison, each of the blocks is assigned the class with the closest centroid, both for the **step 1**-series and the synthetic series. The assigned classes are shown in Figure 4 by the coloured bars at the top, the lower one for the **step 1**-series, the upper one for the synthetic series. If the two classes coincide, the blocks are considered as similar. Otherwise, this block from the **step 1**-series is seen as too distant from the synthetic series and therefore as contributing too much to the mismatch between the regression parameters of the too series. In that case, it is marked for replacement, illustrated in Figure 4 by the shaded rectangles.

[Fig. 4 about here.]

Thereby, on the one hand, blocks which are similar to the synthetic series are kept from the first approximation. This leaves parts of the observed intra-annual sequences unchanged for the simulated series, which contributes to realistic annual cycles in them. On the other hand, blocks which contribute too much to the mismatch between the prescribed regression parameters and those obtained from **step 1**, are replaced. This can finally result in simulated series complying with the prescribed regression parameters.

Step 2 b: Each replacing block is randomly drawn from a set reduced by applying several heuristic criteria. These criteria only allow for blocks which on the one hand bring the regression line of the resulting series closer to the prescribed one and on the other hand make sure that the inserted replacement fits well into the parts of the series which have been set already. The candidate set is defined as follows:

1. Choose all blocks from the observation period which belong to the same class as the block from the synthetic series. These blocks will improve the regression line of the simulated annual means.

2. To avoid too frequent reuse of single blocks: only keep unused blocks in the selection.

3. Keep blocks of which the position within the year lies within a ± 20 days window around the respective position of the block to be replaced. This ensures that only seasonally matching blocks remain in the selection.

4. Keep blocks which connect well with their already chosen predecessor and, if also already defined, their successor.

This fourth criterion identifies valid blocks by comparing the connecting blocks, that is, the second half of the predecessor together with the first half of a candidate block, to the block of the observation series that starts in the middle of the predecessor. If they both belong to the same class, then the candidate is admissible. For an already set successor, this works analogously.

Finally, draw randomly from the resulting set. After all "blanks" are filled that way, the date-to-date mapping of this iteration is fully defined.

Iteration of step 2: Despite the replacements with blocks improving the regression line, the resulting series may still miss the prescribed parameters. For example, consider a series of **step 1** of which the mean is lower than the one of the prescribed regression line. In that case, the blocks which are kept from this **step 1**-series will lead to a systematic negative bias for the mean of the corresponding series, which has to be balanced by the replacing blocks. If their effect turns out to be insufficient, the second step is iterated, where the next iteration uses slightly exaggerated regression parameters for the generation of the synthetic series. These are given by the prescribed parameters plus the difference between given and simulated parameters of the preceding iteration. This results in more blocks which are exchanged and of which the selection is based on a synthetic series that balances the bias of the first approximation more efficiently. These iterations are continued until the prescribed parameters are matched within a certain tolerance.

Both at step 1 and at the end of step 2 b, years or blocks are drawn randomly. This makes any given simulation a stochastic realisation of the population of possible simulations, the range of which can be estimated by generating large ensembles of simulations. The ability of generating large ensembles is a very important feature as it allows the study of extensions of the space of possible climate developments. Dynamical models in general do not offer this possibility, as they are too demanding of computer resources.

2.2 Multi-station simulations

The approach is similar for multi-station simulations, except for that it takes place in a parameter space of higher dimensionality. The ultimate goal however, a date-to-date-mapping, is the same. This means that the series of simulated fields consist of spatial fields that were observed during the observation period. They are spatially consistent thereby, also for tricky variables such as precipitation.

Depending on the station density in the region of interest, it can be advantageous to restrict the number of stations actually considered in the construction of the mapping. This can be a practical necessity (e. g., memory limitations) but also helps to avoid the use of redundant information in the case of stations with highly correlated series. Using parameters like mean, variance and trend estimates for precipitation and temperature from the observation series, cluster analysis can identify climatologically similar stations (see Appendix A). Choosing for each of the resulting classes the station which is closest to the respective centroid, a small set of stations is obtained, which represents the variability among the stations in a generalising way. For each of the representative stations, regression parameters are prescribed, thereby allowing for the simulation of spatially differentiated developments. See Figure 5 for an example from the Elbe river basin.

[Fig. 5 about here.]

For **step 1** – the shuffled annual segments from the observation series – again a large random sample of rearranged annual segments is generated. From this sample, the realisation is chosen for which the regression lines at all of the representative stations are closest to the prescribed ones.

For **step 2**, that is (**2 a**) the identification of blocks to be replaced and (**2 b**) their replacement, works analogously as for the single station case, except for that blocks now are characterised by the characteristic variables of all of the representative stations instead of just one. Consequently, for each of the representative stations a synthetic series is generated, and the blocks of all of these series merged together are compared via the classification to the blocks of the combined series from **step 1**.

If several **iterations** are needed, the exaggerated trends for the next step are calculated as for the single station case for each representative station individually.

2.3 Limitations of the approach

As for all statistical methods, it is necessary to define limits to the applicability of the approach by giving bounds for an appropriate measure of climatological dissimilarity between simulated and observed series. This is not trivial as the study of at least moderate differences between observed and simulated climate are the very reason for its application – long-term means will change, as will annual cycles (Wallace and Osborn 2002).

A plausible criteria therefore focuses on the fluctuations in the series, the noise, which is superimposed on top of long-term trends and annual cycles. If the empirical distributions of the noise of observed and simulated series, respectively, can be seen as characterising the same random variable, then the observation series apparently possesses – despite eventually different trends and annual cycles – enough variability to generate the simulated series for the prescribed regression parameters. The noise signal from a time series can for example be extracted by the STL approach (Cleveland et al. 1990), for the comparison of the two empirical distributions the Kolmogorov-Smirnov test (KS test) is used (see, e. g., von Storch and Zwiers 1999).

3 Performance study: a cross validation

3.1 Experimental setup

To evaluate the quality of the resampling scheme simulations, for the Elbe river basin a period with a known climatology is simulated. Therefore, from the station data of this region covering 1951–2003, two periods of 25 years each are selected, a training period 1951–1975 and a simulation/reference period 1976–2000. Comparing simulated and observed climatology of the reference period gives insight into the ability of the scheme to produce realistic simulated series.

The simulation period 1976–2000 was chosen for three reasons:

1. There is sufficient data independent on the simulation period data for the resampling.

2. For this period, data from a similar experiment conducted using the dynamical RCM REMO (**RE**gional **MO**del, the RCM run at the MPI Hamburg, Germany, see Jacob 2001) is available. This enables a comparison not only between model output and observations but also between different types of models.

3. The observed climatologies of training and reference period differ significantly, both with respect to surface observations such as temperature and with respect to large scale circulation statistics (Gerstengarbe and Werner 2005), that is they result from different atmospheric dynamics. If the resampling reproduces these differences satisfyingly, it is applicable to studies of a changing climate (at least to changes of the same order of magnitude as in this experiment) despite the inherent assumption of stationarity innate to all statistical approaches.

Resampling scheme (RS): Daily data from 32 stations of the Elbe river basin covering 1951–2003 are used, out of which four representative stations are identified as described in Section 2.2. They are shown in Figure 5. Observations include among others several temperature parameters (T_{max}, T_{min} and T_{mean}) and precipitation. They are carefully compiled and homogenised (Österle 2005). The numbering of the stations (as for, e. g., Figure 6) starts at the estuary of the Elbe and increases the closer the station is to the source.

Prescribing the regression parameters estimated from the annual mean temperatures 1976–2000 at the representative stations, the data from the training period is used to generate simulated series for this period.

An ensemble of 1000 simulations is generated, which on a 3 GHz CPU takes approximately 40 hours. The prescribed regression parameters do not yield series violating the limiting criteria proposed in Section 2.3.

Special attention is paid to the evaluation of precipitation: It is well known that the Hellmann measuring devices for precipitation underestimate the true precipitation due to wind, evaporation etc.. As long as only precipitation from the resampling scheme simulations and observations are to be compared, this does not constitute any problem. However, since both of them shall be compared to the REMO simulated precipitations (which simulates true precipiation), an empirical correction is applied to the observed precipitation, which aims at adjusting for the underestimation. This correction is based on Richter (1995) and meta-data of the stations from the German Weather Service (DWD).

REMO: Within an initiative by the German federal environmental agency, climate simulations for Germany 1950–2100 are generated by REMO, and are available to the public (Jacob 2007b). Their horizontal resolution is of 10 km, time step length is 30 s. They are forced by output from the coupled atmosphere-ocean-model ECHAM5/MPI-OM, developed at MPI Hamburg as well, which for the 20th century simulations is constrained by observed radiative forcing (e. g., greenhouse gas concentrations).

The REMO output for 1976–2000 is interpolated to the station locations for the comparison to observational data and the RS simulations. The interpolation is done as recommended in Jacob (2007a).

As stated there, a direct comparison of RCM output to observational data is problematic. First, RCMs produce areal averages rather than point data such as station data, which complicates the comparison. Second, it is known from cross validation experiments that RCMs better reproduce tendencies and variabilities than absolute values. This has to be kept in mind when interpreting the results of the following Section.

3.2 Results

First, the validation considers the entire ensemble of simulations from the resampling scheme and its comparison to observational and REMO data. Second, a single simulation is chosen from the ensemble and analysed in detail.

Statistical analysis and plots are done using the statistical programming language R (R Development Core Team 2004).

3.2.1 Entire Ensemble

Figure 6 shows the long-term mean of temperature. The range of the means estimated from the resampling scheme simulations is very small and agrees satisfyingly with the observed means, deviations being smaller than 0.2 K. The REMO data shows a systematic overestimation of mean temperature, which becomes more pronounced the closer the station is to the source region of the Elbe (the high station numbers) and there exceeds 1 K.

[Fig. 6 about here.]

The corresponding picture for mean annual precipitation is given in Figure 7. As expected, the simulated range from the RS is not as narrow as for temperature, however, for most stations the observed means lie within the range with deviations generally not exceeding 10%. REMO strongly tends to underestimate mean precipitation amounts. A particularly pronounced underestimation occurs for the mountain stations, for example station 15 from the Harz: apparently, the spatial discretisation in the REMO simulations cannot capture orographic effects correctly in these cases.

[Fig. 7 about here.]

Figure 8 shows the standard deviation of daily mean temperatures – again observations and simulations from the RS agree pretty well. The disagreement between REMO and observations is weaker than for the means of temperature and precipitation, however, at least for some of the coastal stations the RS better reproduces the true variability. The stations closer to the coast (corresponding to the small station indexes) show a lower variability than those from the middle and upper estuary. This dampening influence from the sea is known both from observations and GCM simulations, see, e. g., Blender and Fraedrich (2003) and Fraedrich and Blender (2003). This feature is also captured by REMO.

[Fig. 8 about here.]

Exemplarily, for two of the representative stations the simulated annual cycles of temperature are analysed by calculating monthly means (see Figure 9). The simulations by the RS reproduce the observed means very well. Differences between observed means and the median of the simulated means do not exceed 0.5 K. The scattering for the winter months is a little larger than for summer. The agreement between REMO and the observations is good for summer and winter, however, for spring and autumn, large differences occur, which in some cases reach 2 K.

[Fig. 9 about here.]

Of special importance for climate impact studies are extreme events such as dry spells or heavy precipitation. The applicability of any downscaling approach therefore crucially depends on its skills to reproduce such extremes. Figure 10 shows simulated and observed mean dry spell lengths. For the stations close to the coast (indexed with small numbers), a systematic underestimation of the observed dry spell lengths of about half a day becomes apparent for the RS simulations. They are pretty close to the dry spell lengths observed during the training period, however (not shown). As the approach generates series with persistences compatible to the ones of the training series, systematic changes of persistence such as those seen here for the coastal stations cannot be reproduced. REMO also underestimates the true dry spell lengths, however with much larger deviations, which decrease slightly the closer the stations lie to the source.

[Fig. 10 about here.]

The Elbe region in the recent past repeatedly suffered from severe floods, some of them caused by single heavy precipitation events. Future water management is thus in need of reliable estimations of such events. Figure 11 shows mean annual occurrences of daily precipitation exceeding 10 mm. As can be expected for any extreme event, the scattering over the RS simulations is not as narrow as for, e. g., mean temperatures, however, at almost all stations the observed occurrences lie within the range of the RS simulations without any systematic deviation. REMO in contrast underestimates these occurrences strongly.

[Fig. 11 about here.]

3.2.2 Single simulation evaluation

From the entire RS ensemble, one single simulation is selected which represents the median of the ensemble on a scale from wet to dry (see Appendix B). Its time series are compared to the actual observations of the simulation period and to the REMO data by statistical tests for four statistics from the analysis in Section 3.2.1.

1. Mean temperatures of the RS and the REMO simulation each are tested for the Null that they stem from a population with the same mean as the observational data from the simulation period. The test applied is the t-test, taking serial correlation into account as proposed by von Storch and Zwiers (1999).

2. Mean precipitation from observational and simulated data are tested in the same way, using the Mann-Whitney test (see von Storch and Zwiers 1999), which is suitable for non-normal samples. The very little serial correlation in the precipitation series is neglected.

3. Numbers of heavy precipitation events (RR$>$10 mm) per year are examined similarly: The empirical distributions of these annual occurrences according to the RS and the REMO simulations are tested for the Null of being a sample from a population with the same distribution function as the occurrences estimated from the observational data. For this, the two-sample Kolmogorov-Smirnov test is used, which tests whether two samples can be seen as drawn from populations with identical distribution functions.

4. The empirical distributions of dry spell lengths from observational and simulated data are analysed in the same way.

Table 1 summarises the results of the tests: The first column contains the bias (absolute deviation), averaged over all stations, both with respect to the RS and the REMO simulation. For the heavy precipitation days, this refers to the mean annual number, for the dry spells to the mean dry spell. The second column gives the range over all stations of the respective deviation, again for RS and REMO simulations. The third column contains the number of stations, at which the respective test rejects its Null for an error probability of 5%, that is the number of stations at which significant differences between observation and simulations can be detected. For an ideal simulation, all these numbers should vanish. As Table 1 shows, neither the RS nor REMO are able to deliver such a simulation. However, the RS is much closer to this goal than REMO, which agrees with the substantially larger biases between observational and REMO data.

[Table 1 about here.]

4 Conclusions and outlook

This paper introduces a heuristic resampling scheme for regional climate simulations and evaluates its performance in a cross validation experiment, where its simulations are compared both to observational data and to data simulated by a dynamical RCM.

It assembles simulated series from segments of observed series. As the simulated series thereby consist of observations and fields from the observation period, physical consistency both of the simulated fields and of the combinations of different variables is ensured. The only outer constraint to the resampling are the parameters of a regression line, which the annual means of a chosen characteristic variable have to feature. By using several heuristic criteria, the simulated series are assembled such that they possess realistic annual cycles and persistences. Lacking any other constraints, realistic in this context means statistically compatible to the observed series.

For the cross validation experiment, two independent and climatologically different 25 years periods are extracted from a dataset of daily data from the Elbe river basin, one serving as training, the second serving as reference period. The simulations of the climatology of the reference period by the resampling scheme show a remarkable agreement between observed and simulated series, not only for simple statistics such as mean or standard deviation but also for persistence and extreme events. As for the reference period data simulated by the RCM REMO is also available, the same analysis is carried out for the REMO data. It turns out that the resampling scheme clearly outperforms REMO.

A frequent argument in the discussion on advantages and drawbacks of dynamical and statistical downscaling approaches states that statistics are of limited use when a changing climate is the subject of study. This is certainly true, and the here introduced scheme shares this limitation. However, as the example presented in this paper shows, such a limitation may weigh less than the difficulties which RCMs have to face (too coarse spatial and temporal resolution, inaccurate representation of physical processes, incomplete boundary conditions etc.).

RCMs are the tools to be chosen for simulations of climate conditions strongly differing from the present climate and for all kinds of analysis which focus on the physical processes. Statistical downscaling techniques on the other hand are still indispensable when accurate estimations of a near future climate are needed, such as for regional impact studies for the next few decades.

As it is, the scheme presented in this paper provides a fast and easy-to-use tool, which, unlike many of its alternatives, is relatively independent on complex driving information, e. g., extensive GCM output. As the simulation series consist of station observations, plausible data with respect to every aspect of local weather conditions are generated – a feature which any kind of dynamically generated grid box series lacks. Series simulated by this approach may therefore serve as a sound starting point for regional impact studies.

Acknowledgements

Two anonymous reviewers are gratefully acknowledged for their helpful comments on the first version of this manuscript.

A Cluster analysis

Cluster analysis is a classical tool for unsupervised pattern recognition from a set of observations (e. g., Hartigan 1975, Steinhausen and Langer 1977). For this paper, a combination of hierarchical and non-hierarchical cluster analysis is used.

Hierarchical methods, as opposed to their non-hierarchical relatives, do not operate on the observations themselves but on their mutual similarities or distances. The therefore necessary distance matrix of the observations is calculated using the Euclidean distance. Cluster aggregation is based on the Minimal-Variance-distance, which tends to produce hyper-spherical clusters. A thus obtained classification is passed as initial classification to the non-hierarchical k-means algorithm (Hartigan and Wong 1979).

In contrast to non-hierarchical methods, the results from a hierarchical classification do not depend on the initialisation and the order in which the observations are processed. The initial classification it generates for the non-hierarchical part is therefore robust. The non-hierarchical part merely serves as a fine-tuning to achieve the unique representability of the clusters by their centers of mass, a mandatory feature for this application, which by means of a hierarchical cluster analysis alone cannot be guaranteed.

The choice of the cluster number remains to a certain degree subjective. For the number of representative stations, it is based on visual inspection and the climatological knowledge of the region. For the classification of the blocks, for the Elbe data typically 40 clusters are used. This order of magnitude can be found in related literature (e. g. Enke and Spekat 1997). Experiments with differing cluster numbers between 20 and 100 did not yield detectable differences, however (not shown).

B Selection of a single simulation

For the single simulation analysis, the simulation is chosen from the ensemble which on a scale from wet to dry corresponds to the median of the ensemble. The measure used for this is the *climatological water balance* (CWB), that is the difference between precipitation and potential evaporation, which is calculated according to the empirical Turc-Ivanov formula given in DVWK (1996). The CWB thus measures the amount of water which is potentially exchanged between atmosphere and soil, negative values indicating a loss for the soil.

Each simulation is characterised by mean and trend of the annual CWB sums at the representative stations, which gives an eight-tuple (four stations, two parameters each) for each simulation. After calculating the median of each of these eight parameters, the simulation closest to this median-tuple is chosen.

References

Beersma J, Buishand T (2003) Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation. Climate Research 25:121–133

Benestad R (2004) Tentative probabilistic temperature scenarios for northern Europe. Tellus 56A:89–101

Blender R, Fraedrich K (2003) Long time memory in global warming simulations. Geophysical Research Letters 30:1769–1772

Cleveland R, Cleveland W, McRae J, Terpenning I (1990) STL: A Seasonal-Trend Decomposition Procedure Based on Loess. Journal of Official Statistics 6:3–33

Déqué M (2004) Uncertainties in PRUDENCE simulations: Global high resolution models. Tech. rep., DMI. URL http://prudence.dmi.dk/

DVWK (1996) Ermittlung der Verdunstung von Land- und Wasserflächen. DVWK Merkblätter zur Wasserwirtschaft 238, DVWK

Enke W, Spekat A (1997) Downscaling climate model outputs into local and regional weather elements by classification and regression. Climate Research 8:195–207

Fox-Rabinovitz M, Côté J, Dugas B, Déqué M, McGregor J (2006) Variable resolution general circulation models: Streched-grid model intercomparison project (SGMIP). Journal of Geophysical Research 111:1–21

Fraedrich K, Blender R (2003) Scaling of Atmosphere and Ocean Temperature Correlations in Observations and Climate Models. Physical Review Letters 90:108501(1–4)

Gerstengarbe FW, Werner P (2005) Katalog der Großwetterlagen Europas (1881–2004) nach Paul Hess und Helmut Brezowsky, 6. verbesserte und ergänzte Auflage. PIK Report 100, Potsdam Institute for Climate Impact Research

Hartigan J (1975) Clustering Algorithms. Wiley, New York, USA

Hartigan JA, Wong MA (1979) A K-means clustering algorithm. Applied Statistics 28:100–108

Houghton J, Ding Y, Griggs D, Noguer M, van der Linden P, Dai X, Maskell K, Johnson C (2001) Climate Change 2001, The Scientific Bases. Contribution of Working Group I to the Third Assessment Report of the IPCC. Cambridge University Press, Cambridge

Jacob D (2001) A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin. Meteorol Atmos Phys 77:61–73

Jacob D (2007a) Hinweise für REMO-Datennutzer. URL http://www.mad.zmaw.de/fileadmin/extern/documents/REMO-UBA-Hinweise.pdf

Jacob D (2007b) The REMO UBA project. URL http://www.mad.zmaw.de/projects-at-md/sg-adaptation/links-to-other-regional-model-data/remouba-project/

Mitchell J, Ewins P (2003) The Scientific and Technical Review 2002/3. Tech. rep., Met Office, Exceter, UK

Murphy J (1999) An Evaluation of Statistical and Dynamical Techniques for Downscaling Local Climate. Journal of Climate 12:2256–2284

Murphy J, Sexton D, Barnett D, Jones G, Webb M, Collins M, Stainforth D (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430:768–772

Österle H (2005) Annual mean temperatures of Germany according to ECHAM4 present climate simulation. Personal Communication at Potsdam Institute for Climate Impact Research, Germany

R Development Core Team (2004) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org, iSBN 3-900051-00-3

Richter D (1995) Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmann- Niederschlagsmesser. Bericht des DWD 194, Deutscher Wetterdienst, Offenbach, Germany

Ruiz-Barradas A, Nigam S (2006) IPCC's Twentieth-Century Climate Simulations: Varied Representations of North American Hydroclimate Varaibility. Journal of Climate 19:4041–4058

Semenov M, Barrow EM (1997) Use of a stochastic Weather Generator in the Development of Climate Change Scenarios. Climatic Change 35:397–414

Steinhausen D, Langer K (1977) Clusteranalyse – Einführung in Methoden und Verfahren der automatischen Klassifikation. Walter de Gruyter, Berlin

Timbal B, McAvaney B (2001) An analogue-based method to downscale surface air temperature: Application for Australia. Climate Dynamics 17:947–963

von Storch H, Hewitson B, Mearns L (2000) Review of Empirical Downscaling Techniques. In: Regional climate development under global warming (eds. Iversen T, Hoiskar B). No. 4 in General Technical Report, RegClim, Jevnaker, Torbjornrud, Norway, pp. 29–46

von Storch H, Zwiers F (1999) Statistical Analysis in Climate Research. Cambridge University Press, Cambridge, UK

Wallace C, Osborn T (2002) Recent and future modulation of the annual cycle. Climate Research 22:1–11

Wessel P, Smith W (2006) The Generic Mapping Tools – GMT. University of Hawai'i at Manoa, NOAA. URL http://gmt.soest.hawaii.edu/

Wilby R, Wigley T (1997) Downscaling general circulation model output: a review of methods and limitations. Progress in Physical Geography 21:530–548

Wilby R, Wigley T, Conway D, Jones P, Hewitson B, Main J, Wilks D (1998) Statistical downscaling of general circulation model output: A comparison of methods. Water Resources Research 34:2995–3008

Wilks D (1999) Interannual variability and extrem-value characterisitcs of several stochastic daily precipitation models. Water Resources Research 34:2995–3008

Zorita E, von Storch H (1999) The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. Journal of Climate 12:2474–2489

Mailing Address:

Potsdam Institute for Climate Impact Research
P.O. Box 60 12 03
14412 Potsdam
Germany

List of Figures

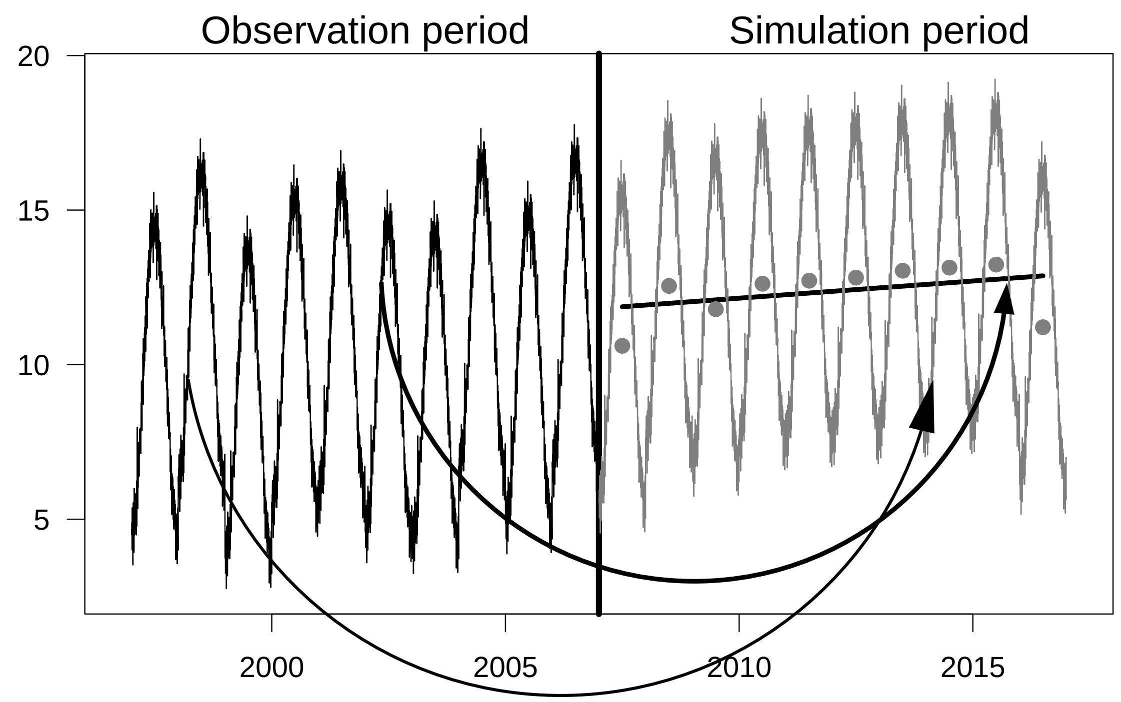

Fig. 1: Assembling simulated series from segments of observed series, corresponding to a prescribed regression line. **Black:** illustrative temperature series from the observation period and the prescribed regression line for the simulation period. **Grey:** simulated temperatures of which the annual means (grey dots) yield the prescribed regression line. This series is assembled from elements of the observation series, as illustrated by the two exemplary arrows.

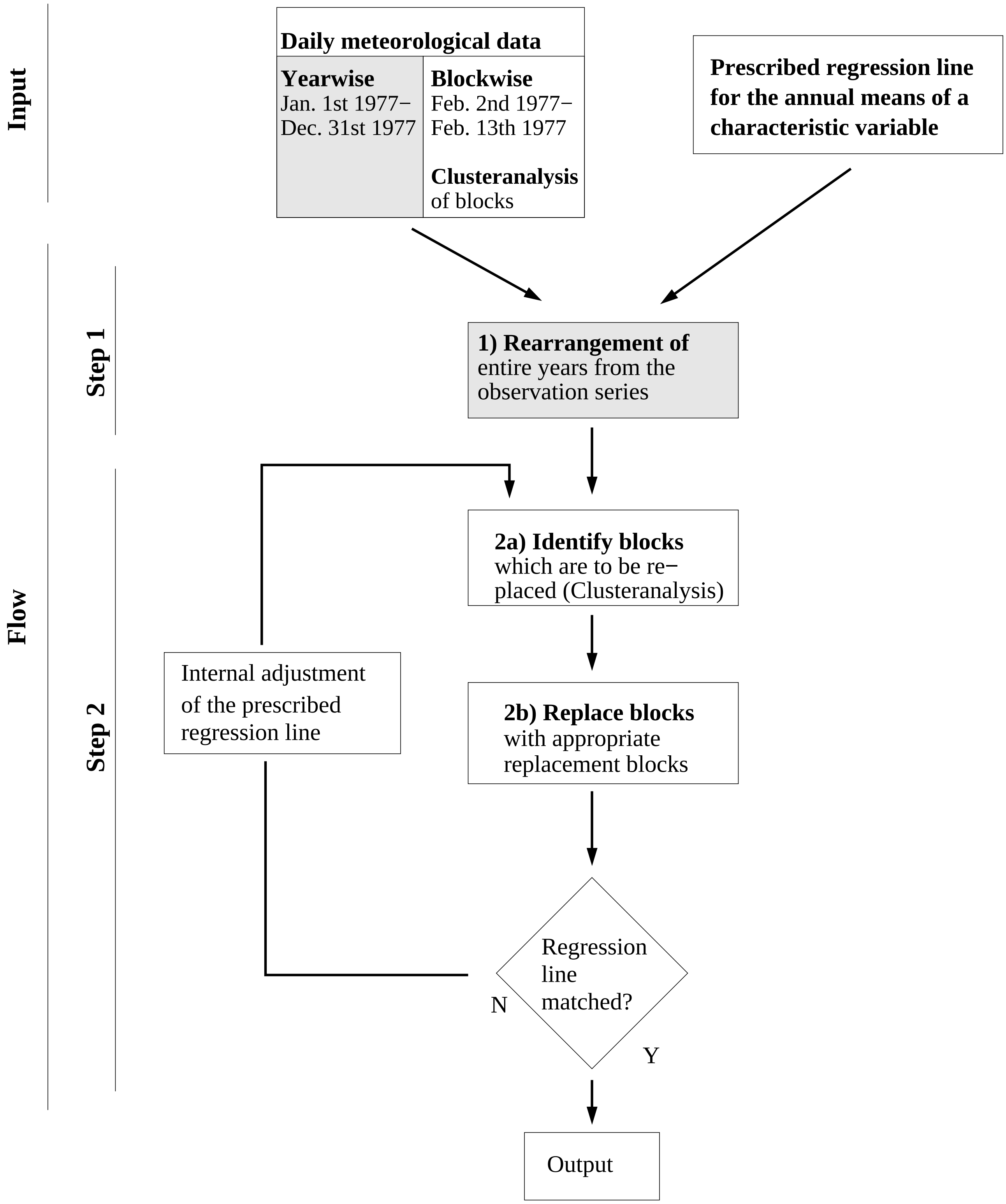

Fig. 2: Summary of the generation of the date-to-date-mapping.

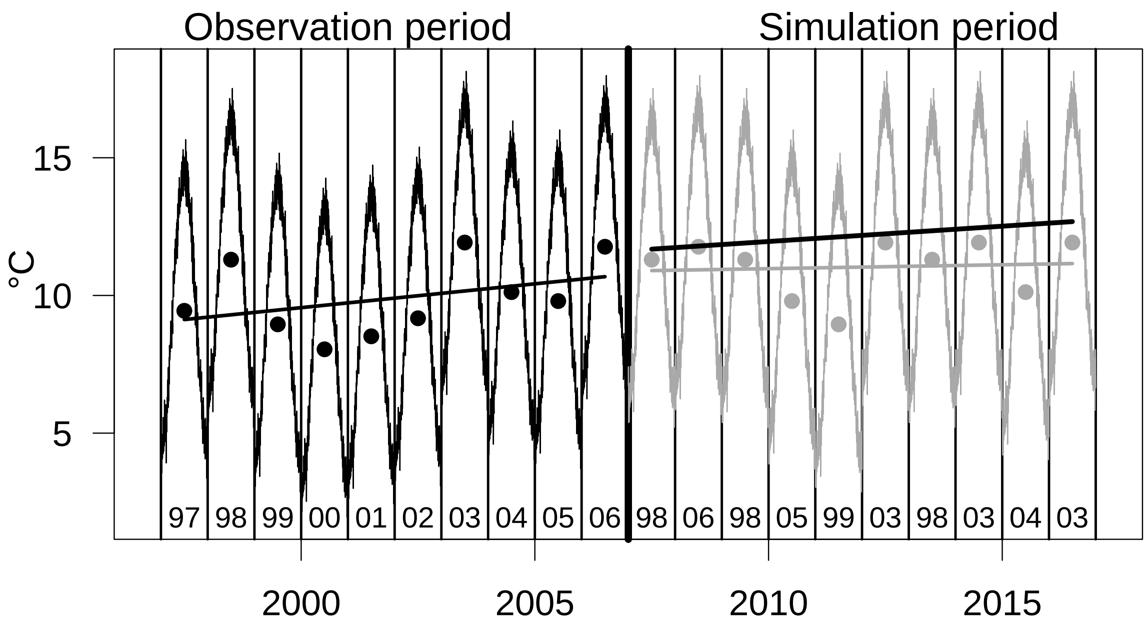

Fig. 3: Step 1: Generating a "first approximation" by randomly rearranging entire years from the observation period (shown is temperature, grey curve). This rearrangement is chosen such that its annual mean temperatures (grey dots) yield a regression line which is close to the prescribed line (black).

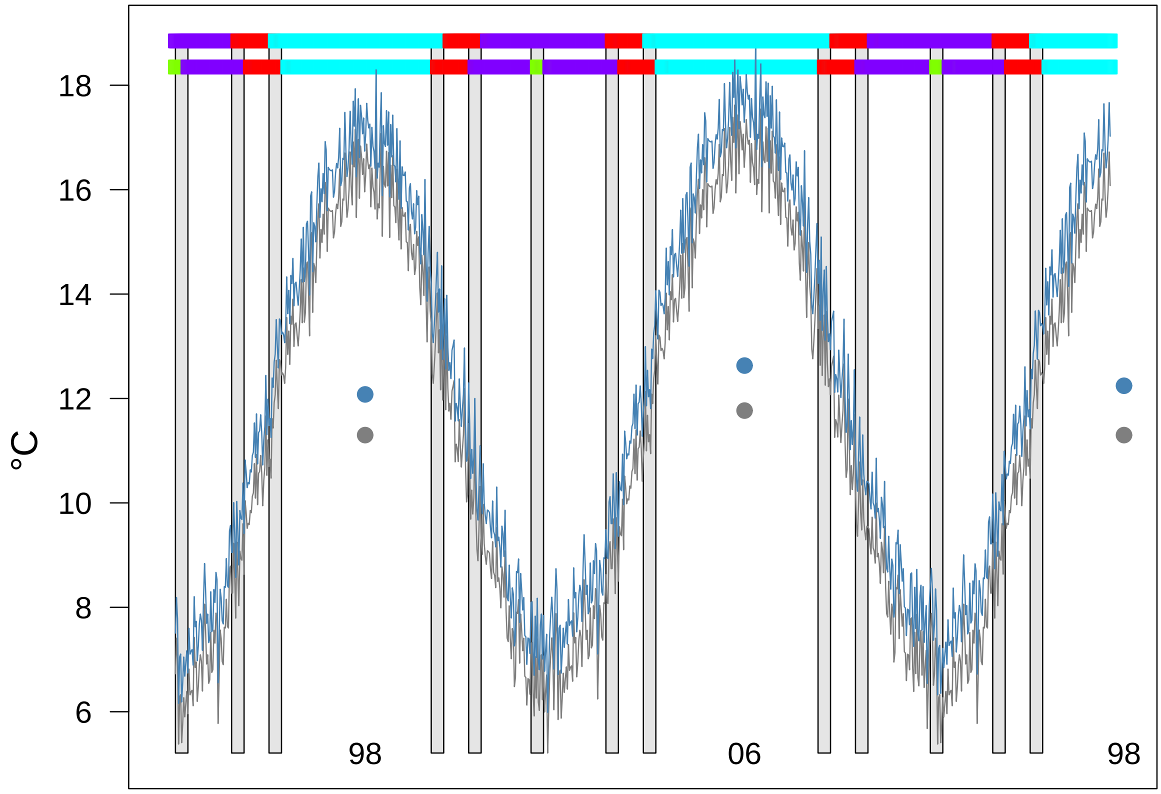

Fig. 4: Step 2 a: The temperature series from **step 1** (grey) and an synthetic series (blue), of which the annual means (dots) match the prescribed regression lines. At the top: assigned classes of each of the blocks with respect to the two series. Where they coincide, the block is kept from **step 1**. Otherwise, it will be replaced (these blocks are marked by the shaded rectangles).

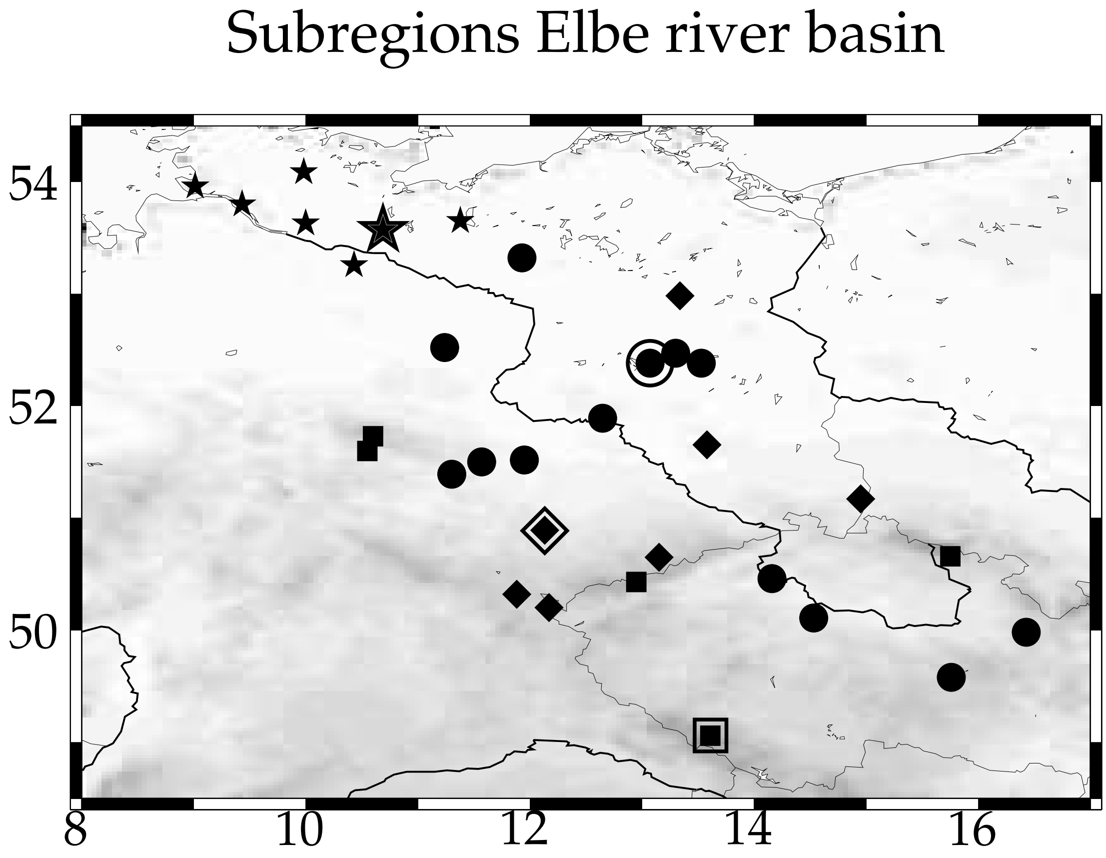

Fig. 5: Stations from the Elbe river basin, with different symbols according to their classes defined by a cluster analysis. The representative stations of each class have their symbols contoured.

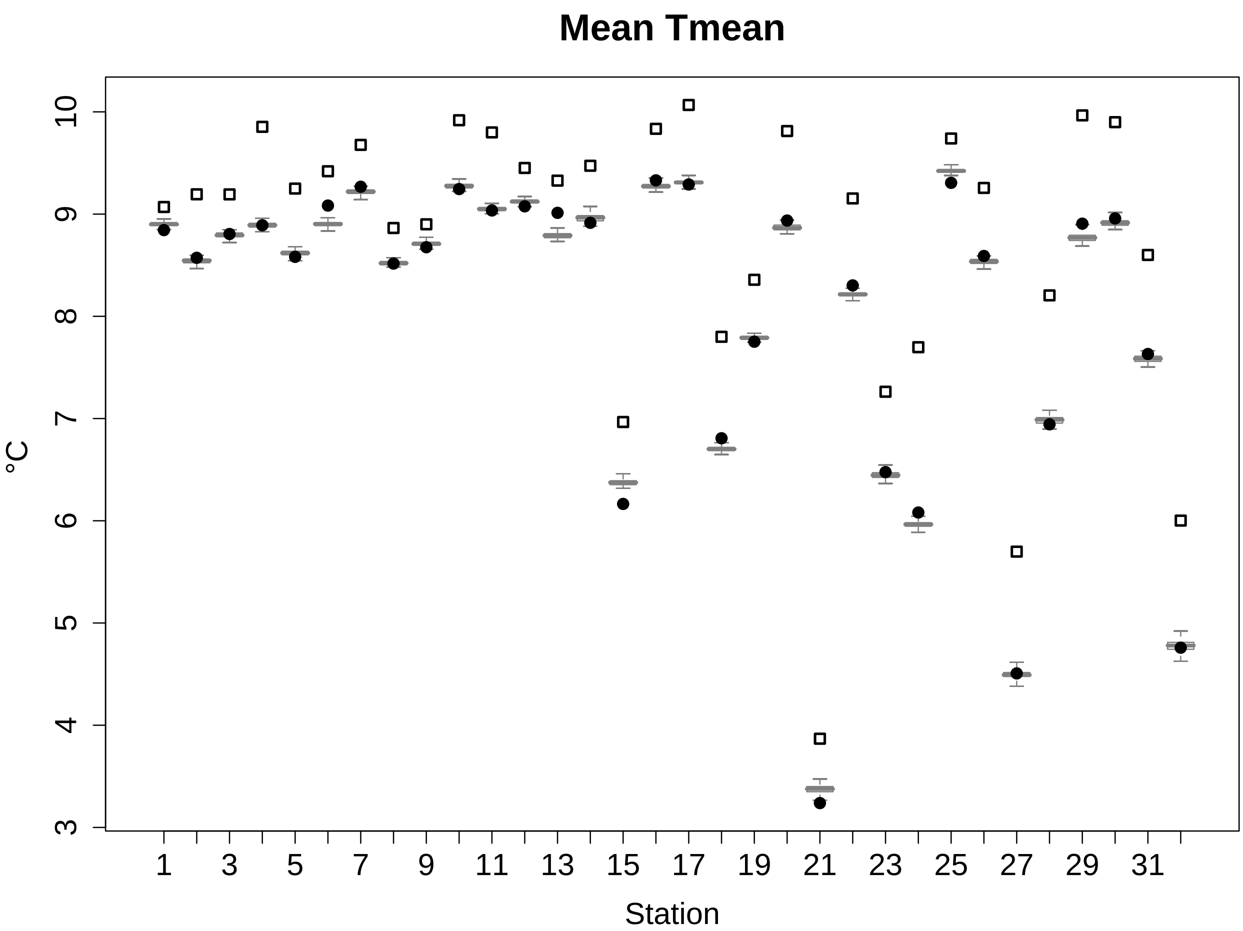

Fig. 6: Long-term mean of temperature 1976–2000. Along the horizontal axis, the 32 stations are numbered. Black dots: observed means 1976–2000. Black squares: Means simulated by REMO. Grey boxplots: distribution of the means across the 1000 RS simulations. The bold line shows its median, the box itself the interquartile range and the whiskers comprise the entire range including outliers.

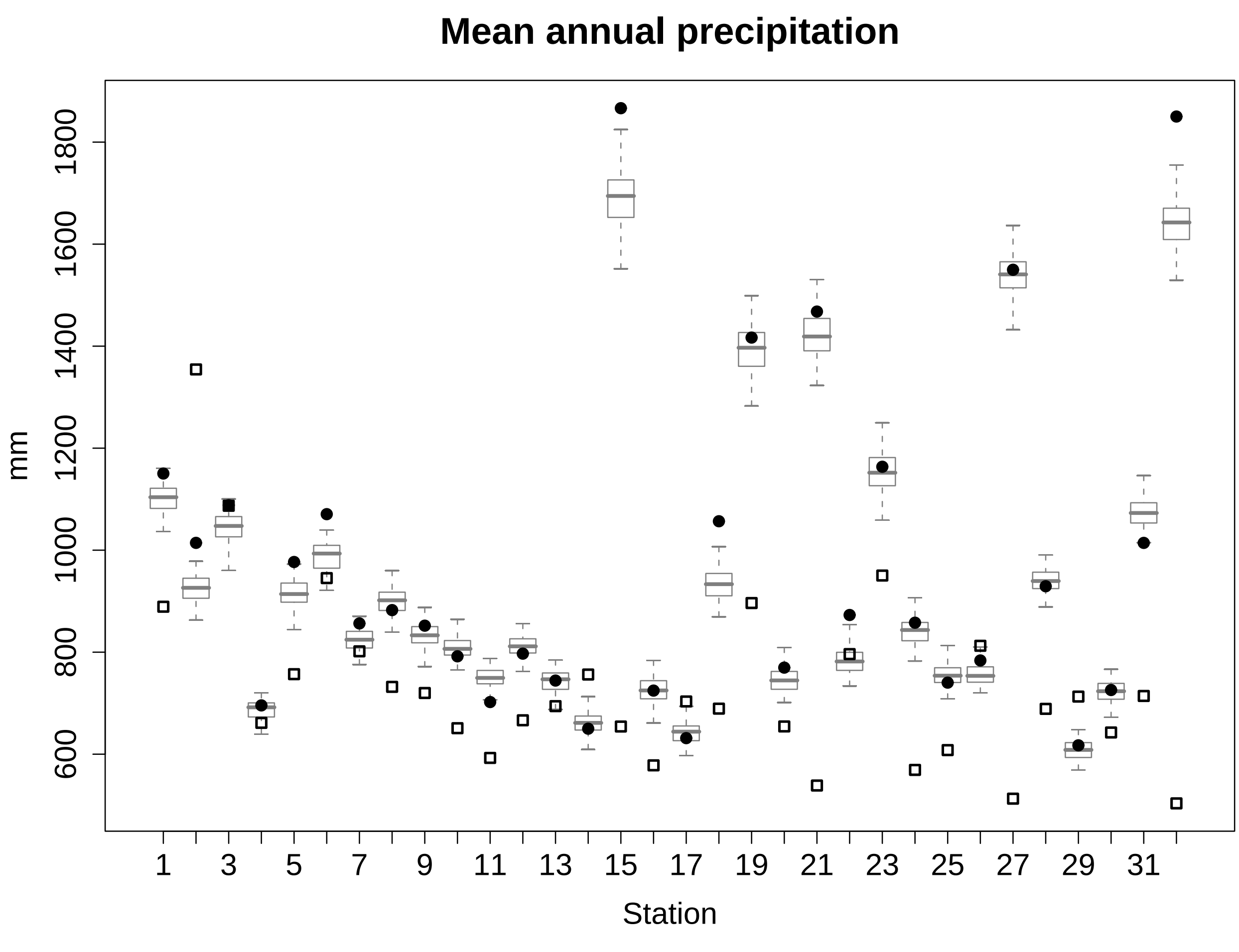

Fig. 7: Long-term mean annual precipitation 1976–2000. The plot is organised as described for Figure 6.

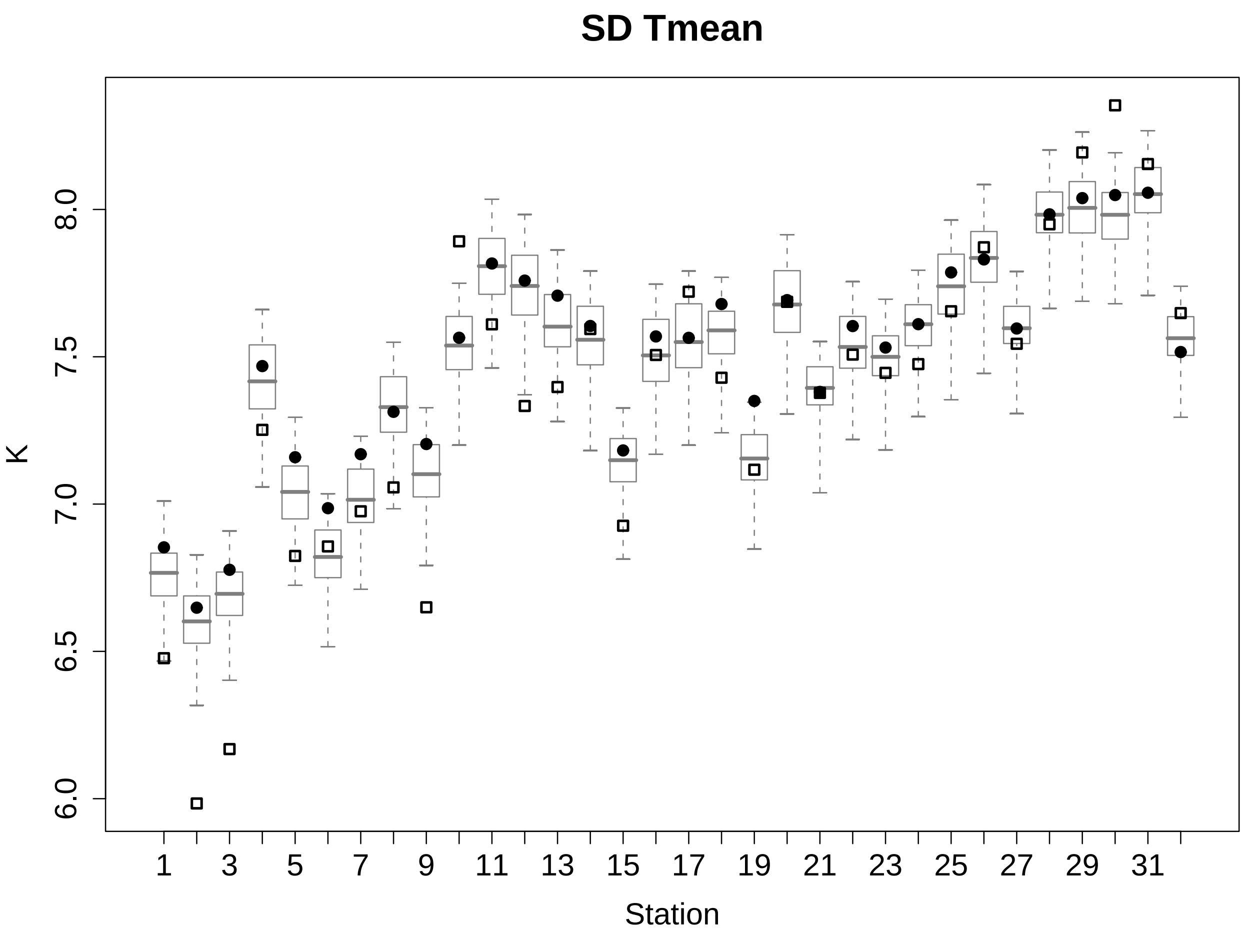

Fig. 8: Long-term standard deviation of temperature 1976–2000. The plot is organised as described for Figure 6.

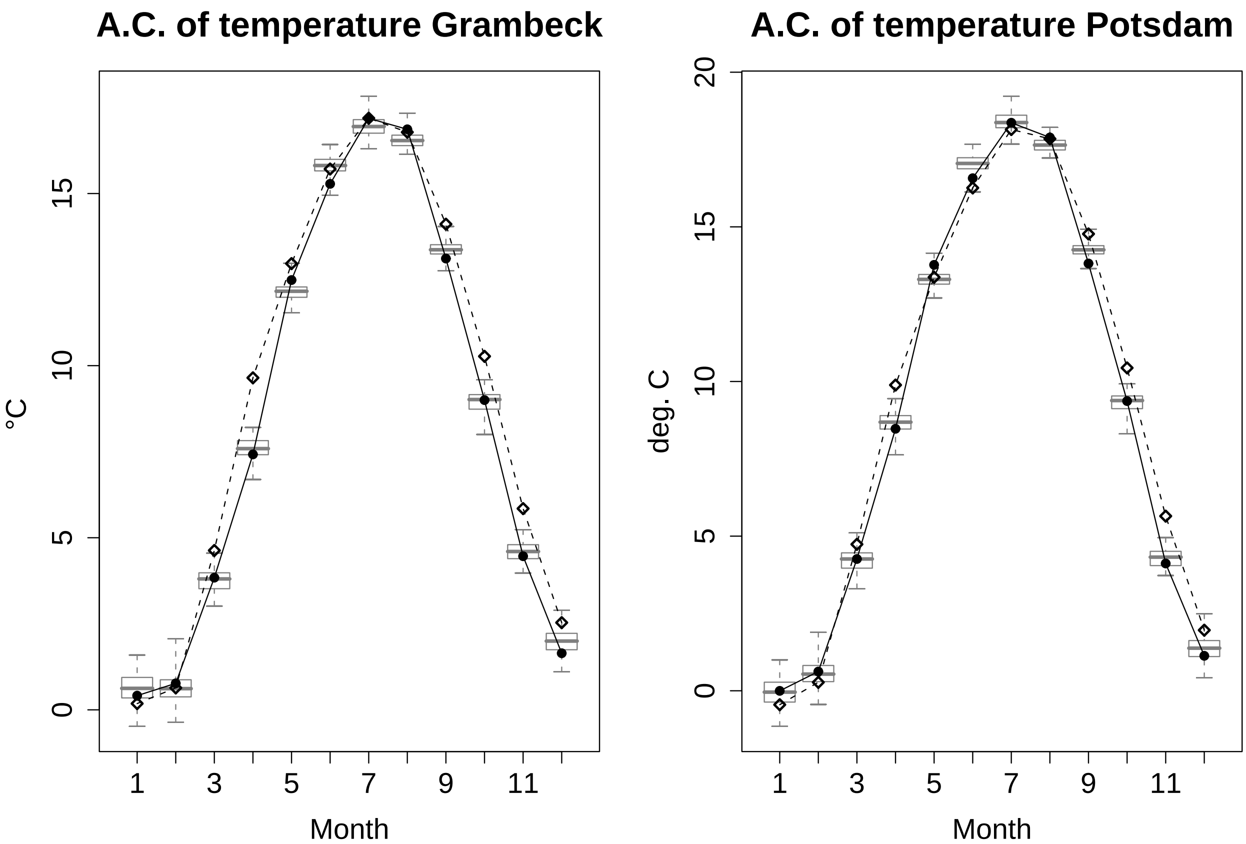

Fig. 9: Long-term monthly means of temperature 1976–2000. Black dots, connected by a solid line: long term monthly means from the observations. Black diamonds, connected by a dashed line: the same for the REMO data. Grey boxplots: the range of the long term monthly means estimated from the 1000 RS simulations (compare to the caption of Figure 6).

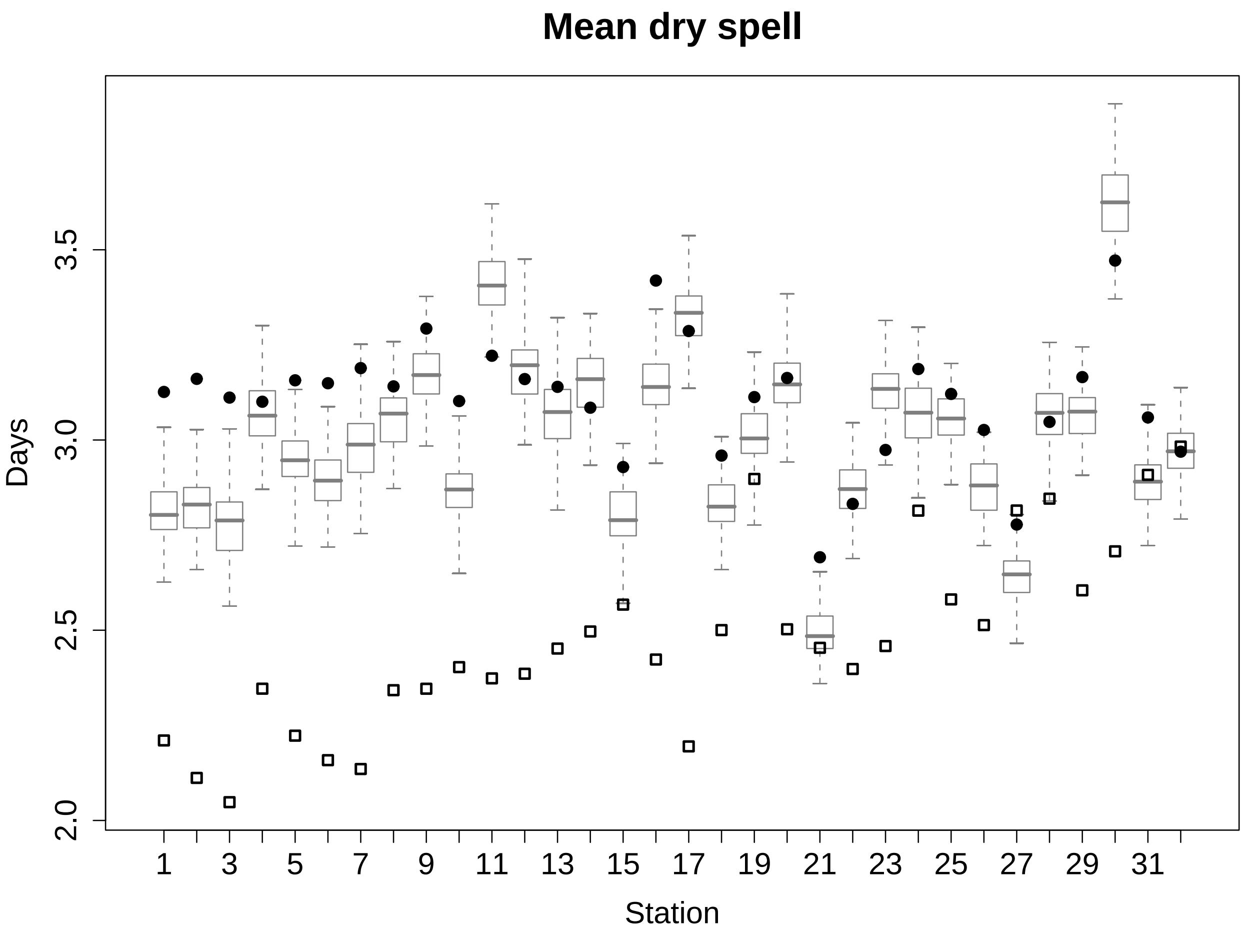

Fig. 10: Mean dry spell 1976–2000. The plot is organised as described for Figure 6.

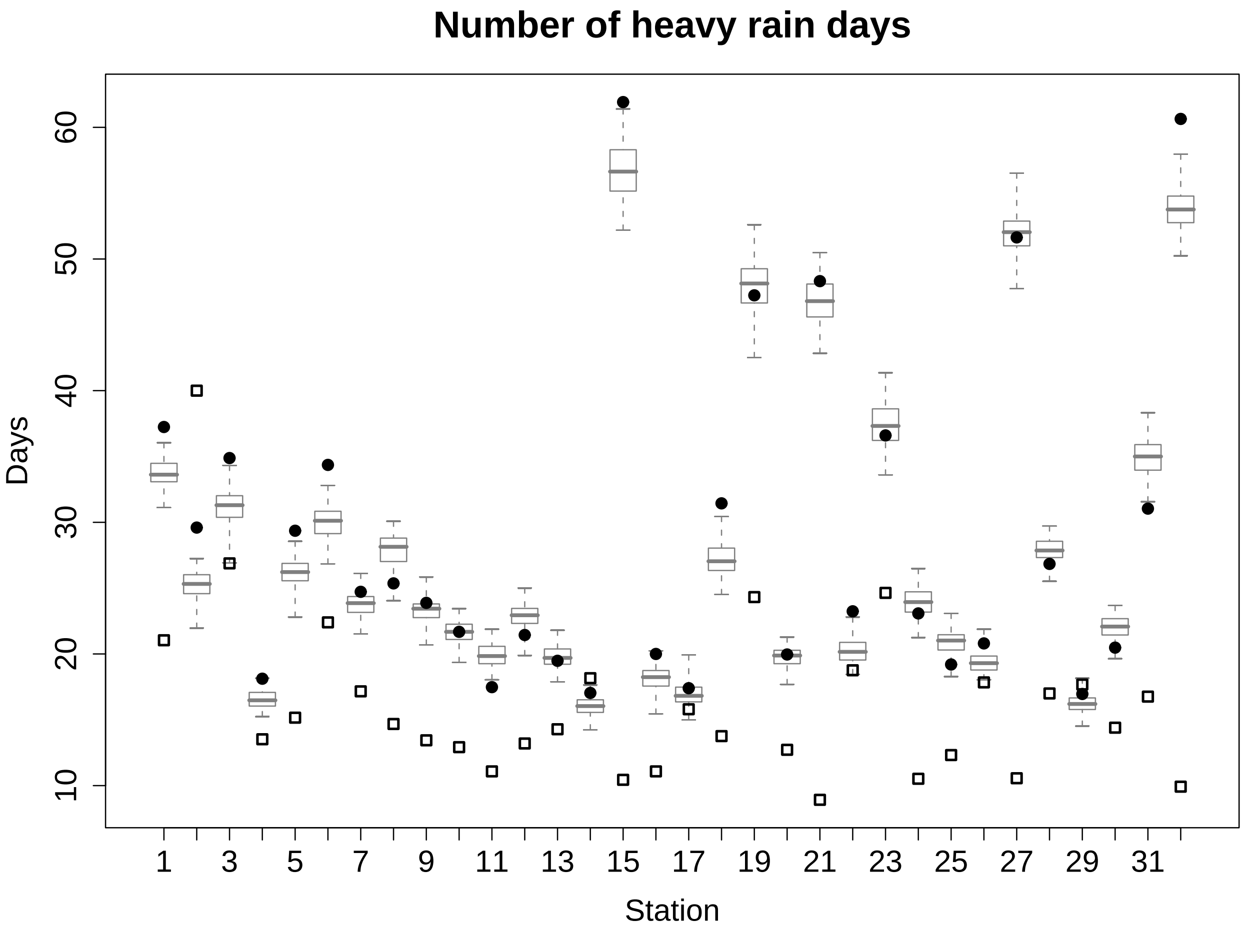

Fig. 11: Mean annual occurrences of daily precipitation $> 10\,\mathrm{mm}$ 1976–2000. The plot is organised as described for Figure 6.

List of Tables

Table 1: Single simulation comparison. Four statistics are evaluated: mean temperature (t-test), mean precipitation (Mann-Whitney-test), annual occurrences of heavy precipitation events and dry spell lengths (Kolmogorov-Smirnov-test for both). Columns give deviations of the respective statistics between simulations and observations and summaries the results from the tests.

	Mean bias		Range deviations		# stations $\mid p < 5\%$	
	RS	REMO	RS	REMO	RS	REMO
T_{mean} (K)	0.07	0.72	-0.17..0.17	0.22..1.62	0	32
RR_{mean} (%)	3.45	27.05	-9.57..4.59	-61.66..81.64	12	32
# Days $\mid$ RR $>$ 10mm	1.23	6.71	-3.96..2.24	-33.16..22.24	1	13
Dry spell (Days)	0.13	0.63	-0.29..0.18	-1.09..0.04	4	30