HAL
open science

Reliability of regional climate model simulations of extremes and of long-term climate

U. Böhm, M. Kücken, D. Hauffe, F.-W. Gerstengarbe, P. C. Werner, M. Flechsig, K. Keuler, A. Block, W. Ahrens, Th. Nocke

To cite this version:

U. Böhm, M. Kücken, D. Hauffe, F.-W. Gerstengarbe, P. C. Werner, et al.. Reliability of regional climate model simulations of extremes and of long-term climate. Natural Hazards and Earth System Sciences, 2004, 4 (3), pp.417-431. hal-00301611

HAL Id: hal-00301611

https://hal.science/hal-00301611v1

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Natural Hazards and Earth System Sciences (2004) 4: 417–431
SRef-ID: 1684-9981/nhess/2004-4-417
© European Geosciences Union 2004

Natural Hazards and Earth System Sciences

Reliability of regional climate model simulations of extremes and of long-term climate

U. Böhm[1,2], M. Kücken[1], D. Hauffe[1], F.-W. Gerstengarbe[1], P. C. Werner[1], M. Flechsig[1], K. Keuler[3], A. Block[3], W. Ahrens[3], and Th. Nocke[4]

[1]Potsdam-Institut für Klimafolgenforschung, Potsdam, Germany
[2]Universität Potsdam, Germany
[3]Brandenburgische Technische Universität Cottbus, Germany
[4]Universität Rostock, Institut für Informatik, Germany

Received: 30 September 2003 – Revised: 15 January 2004 – Accepted: 3 May 2004 – Published: 21 June 2004

Abstract. We present two case studies that demonstrate how a common evaluation methodology can be used to assess the reliability of regional climate model simulations from different fields of research. In Case I, we focused on the agricultural yield loss risk for maize in Northeastern Brazil during a drought linked to an El-Niño event. In Case II, the present-day regional climatic conditions in Europe for a 10-year period are simulated. To comprehensively evaluate the model results for both kinds of investigations, we developed a general methodology. On its basis, we elaborated and implemented modules to assess the quality of model results using both advanced visualization techniques and statistical algorithms. Besides univariate approaches for individual near-surface parameters, we used multivariate statistics to investigate multiple near-surface parameters of interest together. For the latter case, we defined generalized quality measures to quantify the model's accuracy. Furthermore, we elaborated a diagnosis tool applicable for atmospheric variables to assess the model's accuracy in representing the physical processes above the surface under various aspects. By means of this evaluation approach, it could be demonstrated in Case Study I that the accuracy of the applied regional climate model resides at the same level as that we found for another regional model and a global model. Excessive precipitation during the rainy season in coastal regions could be identified as a major contribution leading to this result. In Case Study II, we also identified the accuracy of the investigated mean characteristics for near-surface temperature and precipitation to be comparable to another regional model. In this case, an artificial modulation of the used initial and boundary data during preprocessing could be identified as the major source of error in the simulation. Altogether, the achieved results for the presented investigations indicate the potential of our methodology to be applied as a common test bed to different fields of research in regional climate modeling.

Correspondence to: U. Böhm
(boehm@pik-potsdam.de)

1 Introduction

Quantification of uncertainty is a key issue in the provision of future climate scenarios. In addition to the mean conditions, extremes are an important issue and are often linked to economic damage and loss of property, as during the 2002 Elbe flood or during the unusually dry and hot summer in the Mediterranean region in 2003. Before generating future climate scenarios, a validation of present-day simulations has to be performed to determine the range of errors for the quantities of interest under recent conditions as a raw estimate of their uncertainty in the future.

As regional climate modeling using dynamic models has evolved over the last decade to a widely applied downscaling approach (IPCC, 2001), a variety of evaluation methods were also developed (Machenhauer et al., 1996; Christensen et al., 1997). Quantitative techniques are increasingly utilized (Jones et al., 1995; Machenhauer et al., 1998). Also, evaluation methods for up to three parameters were developed (Taylor, 2001).

Often, however, more parameters are required to fully describe regional climatic conditions or extremes. Although several regional climate modeling inter-comparisons have already been performed or are still ongoing (Takle et al., 1999; Frei et al., 2003), such rather complex evaluations have not been reported.

Therefore, we developed a strategy for a comprehensive evaluation of a model's performance as described in Sect. 2. In addition to individual near-surface parameters, we evaluate multiple near-surface parameters in combination and define general quality measures for this purpose, but also address the question of whether the model behaves reliably above the surface. We used the methodology presented in Sect. 3 to implement this approach into modularly organized software tools and applied the algorithm to evaluate the results of the same regional climate model for two case studies focusing on different research goals and time scales. In Sect. 4, we briefly describe the model and its setup for the two applications.

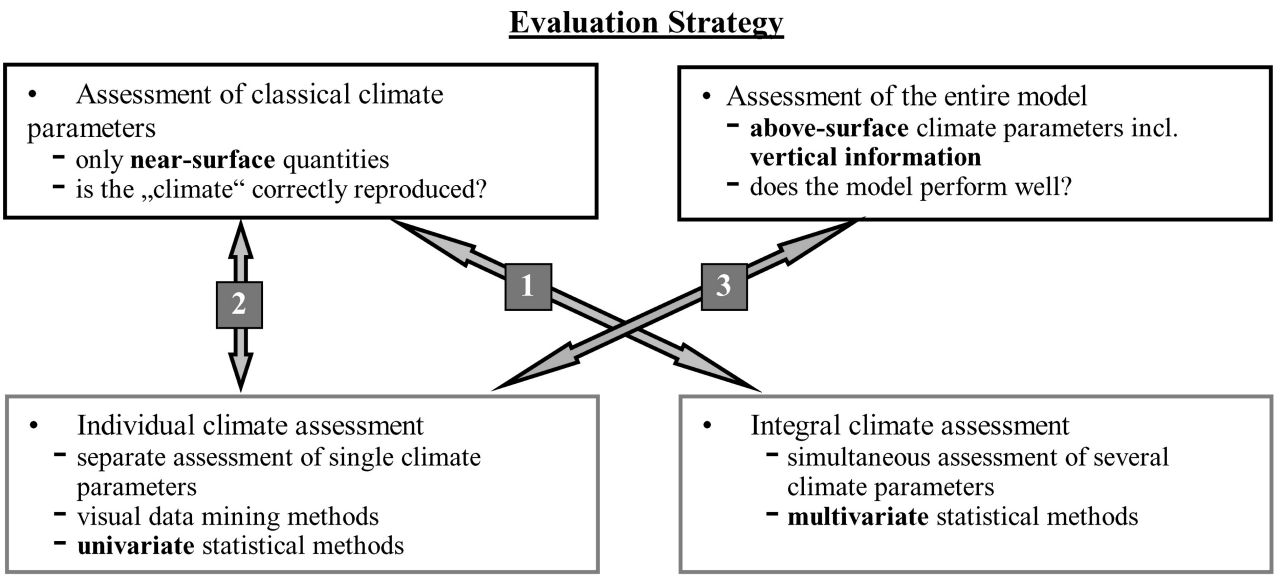

Fig. 1. Basic scheme of the underlying strategy for the developed evaluation methodology.

Section 5 presents the main results obtained by means of the developed evaluation algorithm for Case I focusing on the risk of total yield loss for maize – one of the most important agricultural crops in the Northeast of Brazil – during the drought year 1983. In Sect. 6, we discuss the evaluation results for a long-term present-day regional climate simulation over Europe (Case II) and analyze one of the major sources of deviations observed between model results and reference data.

Multiple model runs such as Monte-Carlo simulations are mainly applicable for the typical time scales of extremes. They provide the opportunity to assess the quality and stability of model results in terms of their statistical properties, for instance for disturbed initial conditions or internal model parameters. We therefore, launched activities to integrate the developed evaluation algorithm into a simulation environment, which allows this kind of experiment. Section 7 gives a brief overview of the SimEnv ensemble simulation environment and the results of a behavioral analysis are discussed as an example. Conclusions are presented in Sect. 8.

2 Evaluation strategy

Model evaluation should be problem-oriented in each case. Only a sub-set of all model variables is of primary interest for any user of the model results, and the required accuracy depends on the actual question to be answered. When regional climate model results are used as input for subsequent impact models, the sensitivity of these models to their input climate parameters may give an estimate of the maximum allowable uncertainty range. Furthermore, the availability and quality of proper reference data affects the minimum possible accuracy of any model evaluation.

Irrespective of these application-specific constraints, each model evaluation should follow some basic guidelines that can be inferred from Fig. 1. In any case, not only the target quantities, but also model variables that represent the state of the underlying physical processes should be assessed (black-rimmed boxes). Both multivariate statistics for an integral appraisal and univariate methods for identifying error contributions from individual model variables and derived quantities (gray-rimmed boxes) should be utilized for each of the two assessment objectives.

In our approach, we combine three different evaluation tracks as shown in Fig. 1 (gray arrows) for a quantitative model evaluation that is as comprehensive as possible.

The first track uses multivariate statistics to examine several near-surface parameters of interest simultaneously. To examine extremes, such sets of parameters may be defined using primary climate variables like 2-m temperature or precipitation, provided they allow the major characteristics of an extreme to be captured. For instance to characterize a dry summer in Europe, parameters such as the number of summer days (daily maximum temperatures $\geq$25°C), the number of hot days (daily maximum temperatures $\geq$30°C), total heat (total daily maximum temperatures $\geq$20°C during the summer), the summer mean of daily mean temperatures and the extreme value mean (mean monthly maximum temperature during the summer) can be utilized. Generalized cost functions allow differences between model results and reference data for all parameters together to be quantified.

In a second track, we diagnose the simulation errors for individual near-surface parameters and/or primary model variables qualitatively from such a kind of representations as difference plots by means of interactive visualization techniques. To quantitatively ascertain these separate error contributions, we apply standard univariate statistics and a definite minimum set of related measures.

The third and final track utilizes univariate statistics to address whether a model performs well and gives a correct representation of the relevant processes in levels above the surface.

We would like to stress that the second and third track of evaluation are predominantly intended to analyze the error contributions for the individual members of a set of parameters that is used beforehand for an integral quality assessment of the model results. In addition, both evaluation tracks are eligible to check the 4-dimensional arrays of prognostic model variables for their deviations from suitable reference data. The causes of these differences, however, may still be obscured in the case of complex model output and highly non-linear error propagation. Sensitivity analyses as outlined in Sect. 7 may help to cope with such problems when the evaluation approach is linked to a multi-run environment in future versions.

The next section presents the methodology we developed to convert the described basic evaluation strategy into applicable software modules for each of the individual tracks. Overall, we combined a multivariate pattern recognition technique with non-parametric statistical tests and developed general error measures for the first evaluation track. For the other two tracks, we used standard univariate statistics together with new defined, problem-related statistical measures.

3 Methodology

For a holistic evaluation of sets of multiple near-surface parameters, we developed software modules for different kinds of data representation and different generalized cost functions with a non-hierarchical cluster analysis algorithm as the basic component. Using this type of pattern recognition algorithm, all members of the total data set to be classified may be redistributed between all clusters in each iteration step of the analysis. Different to this approach, hierarchical methods permit only the subdivision of the members in those clusters, which are found in the previous iteration step, into more clusters, but no exchange between them. For further details, see Steinhausen und Langer (1977).

In particular, we use a minimum distance technique (Forgy, 1965). Gerstengarbe and Werner (1997) and Gerstengarbe et al. (1999) extended this method and included techniques to define an optimized number of initial clusters and for a statistically significant separation of all clusters.

In two different versions, we use this pattern recognition algorithm for both gridded data on the model's grid and for irregularly distributed data at observational station locations.

When extreme periods are investigated, specific criteria dependent on the evaluation goal can be utilized that allow derivation of the parameters describing the extent of the extreme from the original data time series. For Case Study I, we used certain precipitation thresholds for various phases of the growing season (first, second and third month after sowing, between anthesis and grain filling, last month of the growing season and entire growing season). For the evaluation of long-term climate simulations, we used statistical features of the model variables of interest (temporal means and variances) to compile the parameter set to be clustered.

To quantify model uncertainty, we defined a relative error measure between the clusters for model results and reference data at the individual grid points by:

$$R^{\mathrm{norm}}_{\mathrm{Mod,Ref}} = \left(1 - \frac{\nu_{\mathrm{Mod,Ref}}}{R^{\max}}\right) \cdot 100\% \,. \tag{1}$$

with $\nu_{\mathrm{Mod,Ref}}$ expressing the relative overlap between the two clusters for model results and reference data at the grid point of concern and $R^{\max}$ being the maximum possible relative overlap. The details of this approach are described by Kücken et al. (2002).

If the two clusters are identical, $\nu_{\mathrm{Mod,Ref}}$ equals $R^{\max}$ and $R^{\mathrm{norm}}_{\mathrm{Mod,Ref}}$ becomes 0%. If, however, no overlaps exist, $R^{\mathrm{norm}}_{\mathrm{Mod,Ref}}$ reaches its upper limit of 100%. Finally, the average of this quantity for the entire model region gives a general measure of the model's performance. We have used this approach in Case Study II that is described in Sect. 6.

In the version of the developed integral evaluation module that is applicable for irregularly distributed data, model results are first interpolated to the reference station sites. Then, the two sets of parameters for model results and reference data are clustered independently from each other. Therefore, the characteristics of the individual members in both sets of clusters may differ in general. This requires to map both sets of clusters to each other. We used the Euclidean distance between the cluster centroids, the mean value of all normalized parameters in a cluster, to diagnose these relations and to identify for each cluster that is analyzed from the model results the most similar cluster from the reference data set.

For an integral assessment of the differences between model results and reference data, we defined a generalized cost function Q_1 by combining non-parametric test statistics with Euclidean distance measures. As the main idea, we estimate whether the empirical distribution functions for each individual parameter differ for each pair of the most similar clusters from model results and reference data and compute the mean result for all parameters $\overline{Diff}$. This term is multiplied by an expression describing the Euclidean distance between the most similar clusters in the two data sets. To get dimensionless ratios, these distances are related to the mean distance between all clusters from the reference data set $\bar{r}_{\mathrm{Clu,Ref}}$

$$Q_1 = \overline{Diff} \cdot \left\{1 + \frac{1}{\bar{r}_{\mathrm{Ref}}}\left[\alpha \cdot \bar{r}_{\mathrm{Mod,Ref,a}} + (1-\alpha) \cdot \bar{r}_{\mathrm{Mod,Ref,b}}\right]\right\}$$
$$\alpha = \frac{N_{\mathrm{Clu,Ref,a}}}{N_{\mathrm{Clu,Ref}}} \,. \tag{2}$$

Here, $\bar{r}_{\mathrm{Clu,Mod,Ref,a}}$ characterizes the mean distance between all pairs of clusters from model results and reference data. The second term includes the distance measure $\bar{r}_{\mathrm{Clu,Mod,Ref,b}}$ as defined in Böhm (1999). It accounts for cases where some clusters for the reference data could not be assigned as nearest neighbors to any cluster computed for the model results.

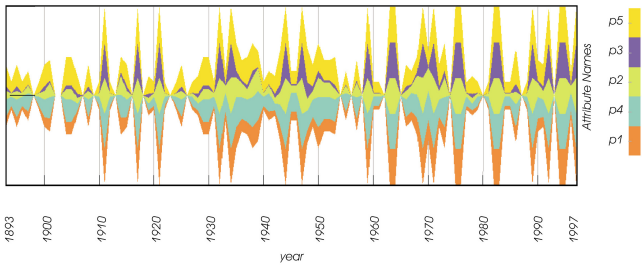

Fig. 2. Example for the use of an adapted ThemeRiver technique (Havre et al., 2002) to diagnose the temporal evolution of five parameters describing extremely hot summers.

Both expressions are complementarily weighted by the ratio α between the number of reference data clusters that are estimated as nearest neighbors for any of model result clusters, and the total number of clusters that are detected for the utilized reference data. To additionally capture structural differences, we defined a quality measure Q_2 that considers deviations between modeled and measured data at the individual locations i of the reference data as a dimensionless ratio:

$$Q_2 = \frac{\frac{1}{N_{\mathrm{Ref}}} \sum_{i=1}^{N_{\mathrm{Ref}}} \overline{r}_{\mathrm{Mod,Ref,i}}}{\overline{r}_{\mathrm{Ref}}} . \tag{3}$$

with the total number of reference stations N_{Ref}. For an ideal model, both quality measures become zero, and their values increase with increasing differences between model results and reference data. For more details, see (Böhm, 1999).

In addition to the separate investigation of individual near-surface climatic quantities, the second track in our evaluation strategy aims also at the diagnosis of the error contributions from the individual members of a parameter set that is used for a preceding holistic evaluation. For both purposes, highly interactive visualization and univariate statistical methods are deployed.

For example, relations between the individual members of a parameter set can qualitatively be diagnosed using suitable visualization techniques. For this purpose, we adapted a simplified theme river representation shown in Fig. 2. In this representation, the temporal evolution of five parameters that are able to characterize hot summers in Europe is displayed for the observation station Potsdam. Low parameter values or a "thin river" period represents cold summers, whereas extremely hot summers are linked to "broad river" episodes. This method makes the clear increase in the number of extreme hot summers during the second part of the 20th century visible. Furthermore, the increasing contributions of parameter p1 (total heat, orange) and parameter p3 (number of hot days, purple) indicate an increase of extreme temperatures rather than longer periods with moderate warm conditions. Further examples for qualitative model evaluation using information visualization methods are described by Nocke et al. (2003).

To quantify individual error contributions, non-parametric statistics are utilized. Although less sensitive if the data follow a normal distribution, the application range is much wider and does not depend on the shape of the real frequency distribution function. Examples of non-parametric statistics are the χ^2 test and the Smirnov test (Taubenheim, 1969), methods to estimate the extreme value range of a distribution (Gerstengarbe and Werner, 1989) or the onset of a trend (Pettitt, 1979) and the four-quadrant test for correlations (Taubenheim, 1969). In addition, basic statistics like bias (signed difference) or root mean square error (e.g. Press et al., 1986) are employed.

In the third track of our evaluation strategy aiming at an assessment of whether the model gives a correct representation of the relevant atmospheric processes, we implemented an open top-down approach to investigate 4-dimensional model variables at levels above the surface under different aspects. Here, top-down describes the degree of detail regarding the spatio-temporal dimensionality of the evaluated arrays. It ranges from "highly aggregated" for time series of area averages at vertical pressure or model levels to "non-aggregated" for 2-dimensional cross sections for all spatial dimensions and at that specific time of the simulated period when the strongest deviations between model results and reference data appear. In addition, we investigated the evolution of the deviations between model results and reference data over the integration period and their statistical characteristics. In the recent version, three diagnosis aspects are considered which shall be mentioned here briefly. First, the bias of area means is used for a general estimate of the model's performance. Second, we focus on the magnitude of the largest differences at individual grid points. Third, structural differences between model results and reference data are addressed. This module is open in the sense that further analysis aspects may easily be added depending on a user's interest.

4 The regional climate model

For our investigations, we developed a climate version (CLM) of the Local Model (LM) of the Deutscher Wetterdienst (DWD). It is a non-hydrostatic limited area atmospheric prediction model. It is designed mainly for operational numerical weather forecast. In the current version 2.19 it is used at a horizontal resolution of about 7 km. The model is formulated on the basis of the primitive thermo-hydrodynamic equations that describe compressible flow in a moist atmosphere. A variety of diabatic sub-grid-scale processes are taken into account by parametrization schemes. For more details, we refer to Doms and Schättler (1999). LM can be applied in a parallel environment and is implemented on an IBM multiprocessor supercomputer. We have added features to the model itself (Kücken and Hauffe, 2002) and to the preprocessing algorithm (Böhm et al., 2002) that makes it possible to supply dynamic lower and lateral boundaries for soil, vegetation and ozone characteristics and to restart the model for long-term simulations.

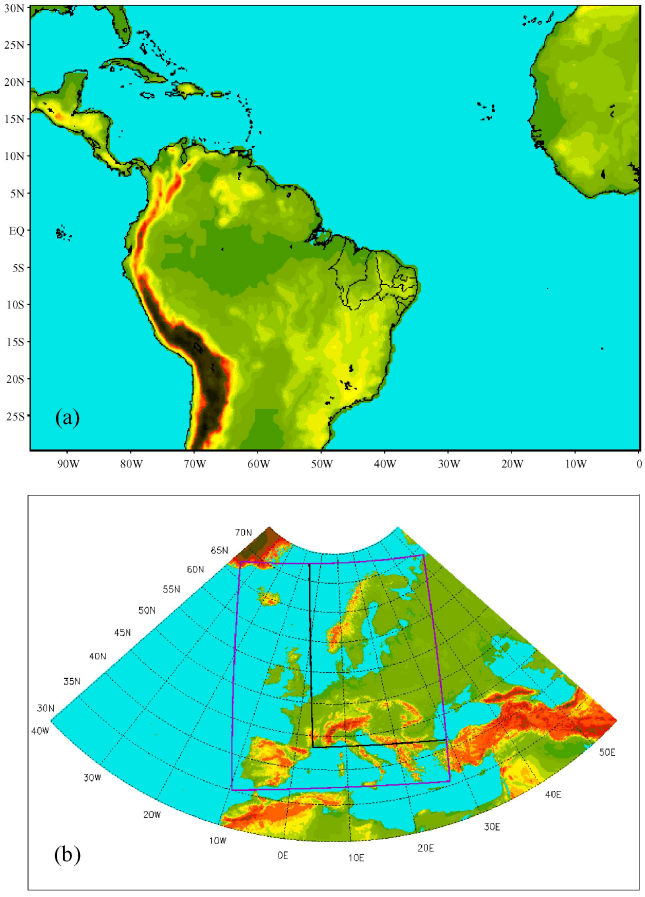

Fig. 3. Model regions for the CLM simulations. Setup **(a)** for the investigation of the agricultural yield risk for the Northeast of Brazil (entire area), **(b)** for long-term simulations over Europe (area bordered by purple lines).

For the investigation of an agricultural drought in the Northeast of Brazil, we setup the model for the area as shown in Fig. 3a. For a long-term climate run, we used the larger of the two regions shown in Fig. 3b. For the two case studies described below, we applied the model with a horizontal resolution of $0.5° \times 0.5°$ in zonal and meridional direction and 20 vertical levels.

5 Agricultural yield loss risk

Case Study I is performed for the extremely dry year 1983 for the Northeast of Brazil. In this year, one of the most severe droughts ever occurred there, linked to drastic yield loss, hunger, impoverishment and migration of people out of the region. We ran the model for this year using initial and lateral boundary conditions that we derived from reanalyses (Gibson et al., 1997) from the European Centre for Medium-Range Weather Forecasts (ECMWF). This investigation should answer the question of whether the regional climate model is able to better reproduce the drought conditions than other regional and global models.

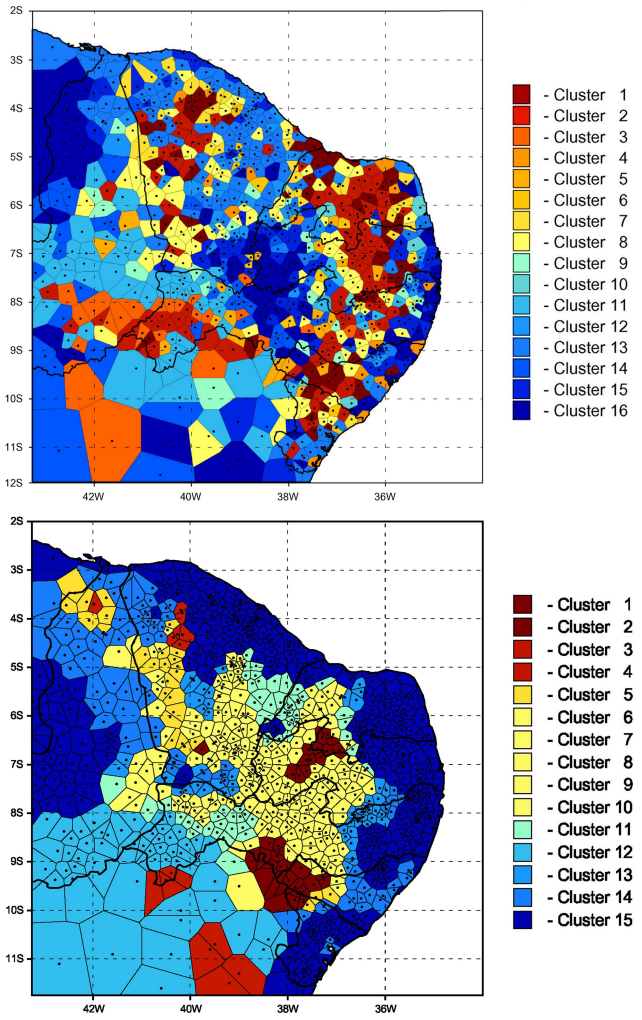

Fig. 4. Cluster analysis results for the risk of potential total yield loss for maize in the Northeast of Brazil 1983. Upper panel: observations, lower panel: model results.

We investigated the agricultural yield loss risk arising from below-normal precipitation during the year 1983 in terms of a potential with idealized soil conditions. Soil scientists from the University of Hohenheim (personal communication) provided us with critical precipitation thresholds for the different phases of the growing season that are mentioned in Sect. 3 for the major agricultural crops in the Northeast of Brazil for these assumptions.

Here, we focused on maize as the most important agricultural crop in the region in terms of the cultivation area for the investigation period. We used monthly observations from 865 stations as reference data and interpolated the model results to the station sites. We derived the corresponding parameters for a cluster analysis from both model result and observational time series using the precipitation threshold criteria in such a way that a parameter is set to zero if rainfall exceeded a threshold, and otherwise it is set to the difference between the threshold value and the rainfall amount. Only if all parameters equal zero, no total yield loss is to be expected.

Table 1. Mean, standard deviation and maximum [mm] for all station sites for the parameters used for the cluster analysis for the risk of total yield loss for maize in the Northeast of Brazil 1983 assuming an idealized growing season (January–April).
Parameters are the difference (threshold value-simulated/observed precipitation) and zero for negative differences.

Parameter	Mean		Standard deviation		Maximum	
	CLM	Observations	CLM	Observations	CLM	Observations
Total precipitation January <60 mm (first month after sowing)	23.65	33.30	19.18	23.04	60	60
Total precipitation February < 70 mm (second month after sowing)	0.53	6.32	2.61	14.29	24	70
Total precipitation March <70 mm (third month after sowing)	0.02	8.13	0.49	15.85	14	70
Total precipitation April <60 mm (last month of the growing season)	23.95	22.14	22.43	21.69	60	60
Total precipitation March and April <130 mm (between anthesis and grain filling)	0.74	20.21	5.30	31.27	65	130
Total precipitation January–April <300 mm (total growing season)	6.62	46.40	21.17	65.06	177	300

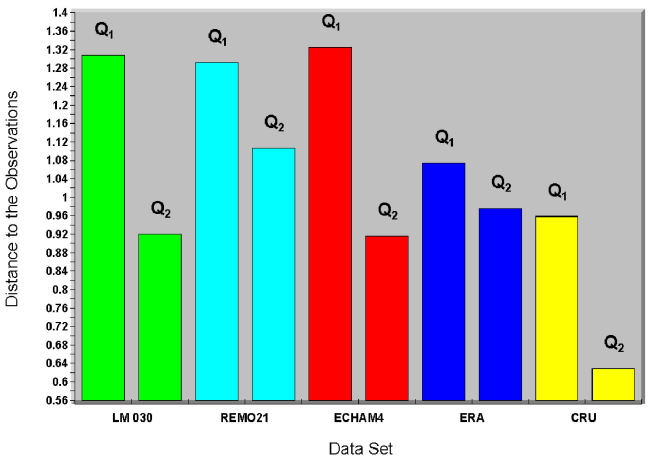

Fig. 5. Generalized quality measures Q_1 and Q_2 for CLM (LM 030), the hydrostatic regional climate model REMO (REMO21), a global climate model (ECHAM4), the driving analyses for the regional models (ERA) and a data set of gridded observations (CRU).

More details are given by Böhm (1999). The cluster analysis was performed for both sets of parameters separately, and the most similar clusters for model results and observational data were identified as described in Sect. 3. For a visual representation, same colors were assigned to these cluster pairs. The resulting patterns are shown in Fig. 4. Red colors indicate a high risk of potential total yield loss for maize, whereas blue colors describe good growing conditions.

At a first glance, it can be seen that a comparable number of clusters was identified for the two data sets. This indicates a comparable level of underlying rainfall variability. But there exist remarkable differences with respect to the spatial distribution of the pairs of the most similar clusters. Clearly, the model underestimates the yield loss risk in most parts of the area investigated, and at the same time it enlarges the most vulnerable regions. We have computed the quality measures Q_1 and Q_2 as described in Sect. 3 for CLM (LM 030) and compared it with a reference run (REMO21) of the regional model REMO (Jacob et al., 1995), a run of the global model ECHAM4 (Roeckner et al., 1996), the driving ECMWF analyses (ERA) and a gridded observation data set (CRU; New et al., 2000).

Without focusing on details of all these data sets, it is visible in Fig. 5 that CLM provides an overall accuracy of the quality measure Q_1that is comparable to that of the other regional model (REMO). This implies a similar magnitude for the deviations between the most similar clusters from model results and observations, as expressed by their differences of the empirical distribution functions for the utilized parameters, and the Euclidean distance between their cluster centroids. Q_2 reaches a smaller value for CLM compared to the other regional model run which can be attributed to a better structural representation of the drought patterns. For details on the intercomparison of CLM and REMO and the major source of error for the latter see Böhm et al. (2003). To improve the performance of CLM as compared to the global model ECHAM4 or the driving ECMWF analyses, however, the ability of the model to simulate precipitation must be enhanced. Another interesting aspect was the relatively high error for the drought intensity in the gridded observations.

An analysis of the individual parameters in terms of their spatial means, standard deviations and maximum values (given in Table 1) indicates that the model simulated too much precipitation for February and, especially, March 1983

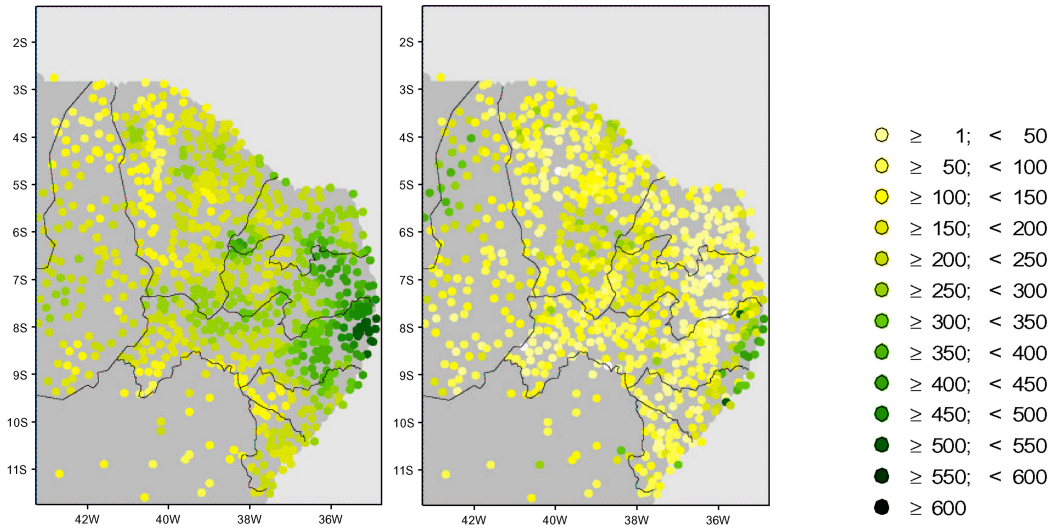

Fig. 6. Total precipitation for March 1983 in the Northeast of Brazil [mm]. Left panel: CLM results, interpolated to station sites, right panel: Observations.

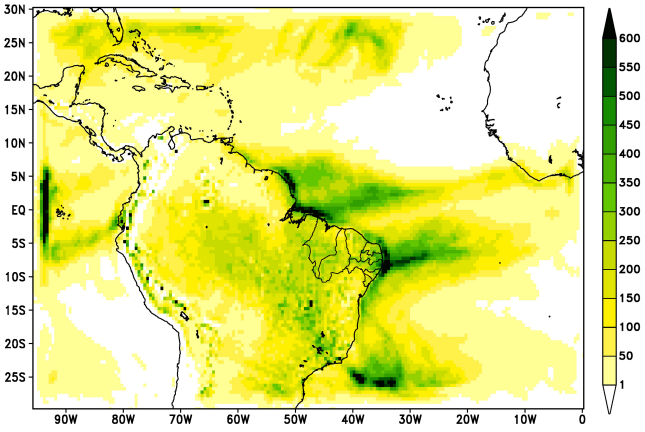

Fig. 7. Total precipitation for March 1983 for the entire model area [mm].

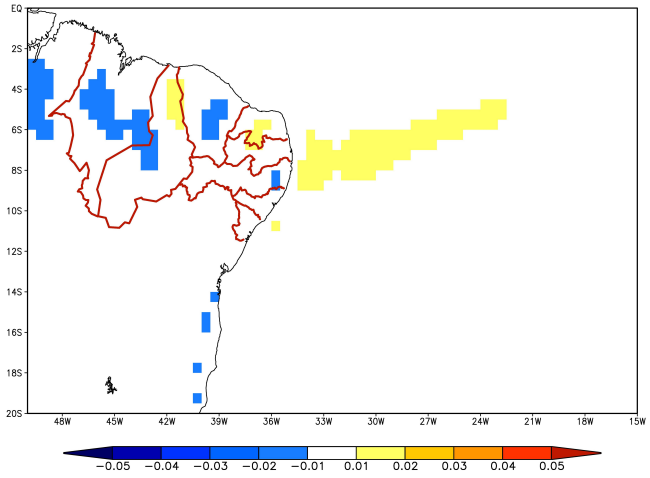

Fig. 8. Monthly mean vertical velocity differences CLM–ECMWF analyses at the level 850 hPa for March 1983 over the equatorial Atlantic region [m/s].

(smaller means for the deficit) with a reduced spatial variability compared to the observations (which has implications also for the two last parameters). The simulated total rainfall in January and April, however, agrees quite well.

The finding that the model simulates too much rainfall for March is confirmed by an evaluation of the spatial patterns for the third parameter (not shown here). At almost all station sites in the Northeast of Brazil, more than 70 mm total precipitation is simulated for March, that is, the mean in Table 1 represents mainly parameter values of 0. Having now identified the parameter that gives the major contributions to the diagnosed deviations between model and observations, we can now concentrate on the underlying model variables and analyze the simulated total rainfall in the diagnosis region for March 1983 that is shown in Fig. 6 together with the reference data. It is obvious that the surplus of rainfall is simulated mostly in regions close to the east coast by the model, whereas in other areas, a better agreement with the reference data is evident. Also, the general spatial patterns of precipitation are rather reasonably captured which explains the good results for Q_2. Thereafter, we investigated the total precipitation for the entire model area for March 1983 as shown in Fig. 7 to get further indications of the sources of this deficiency in the model.

The total area mean of 78 mm is somewhat less than the one computed for analyses from ECMWF (88 mm), but with a relative error of about 11% still comparable. Different to this reference data set, CLM simulated excessive precipitation in the transition zone between sea and land at the Atlantic coast around the equator, where the ITCZ

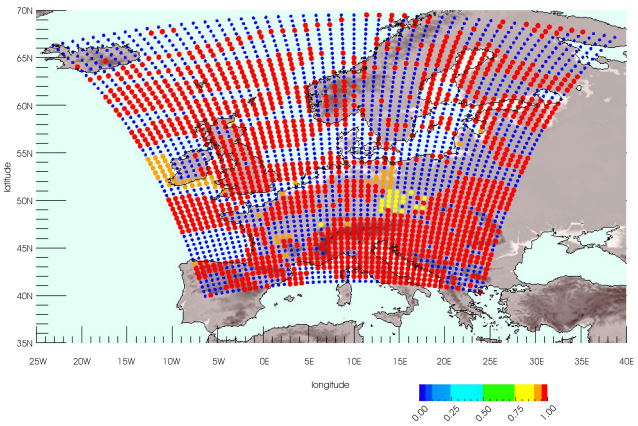

Fig. 9. Patterns of $R^{\mathrm{norm}}_{\mathrm{Mod,Ref}}$ as an integral measure of the differences between the results of a cluster analysis using temperature and specific humidity at the lowest model level for ERA data and CLM model results (run0).

(Intertropical Convergence Zone) is located in March. This surplus of rainfall is, in turn, linked to slightly too high vertical velocities in the model at levels around 850 hPa in this region as indicated in Fig. 8.

Based on these results, we postulated the hypothesis that in coastal regions, when air masses approach the coast of South America, too strong vertical mass fluxes are generated by the dynamic core of the model. To further investigate this problem, we plan to implement a measure for the mass balance into our evaluation algorithm.

6 Present-day regional climate simulation

The second study aimed to reproduce the climate in terms of its statistical characteristics for the 10-year period 1979–1988 over Europe. As for Case I, we supplied initial and lateral boundary conditions from ECMWF re-analyses (ERA) every six hours to the model. This first reference model run is subsequently labeled as CLM run0. The central questions for this research task were (i) whether a non-hydrostatic model version behaves in a numerically stable way for such a large region as indicated in Fig. 3b when it is equipped only with the extensions as described in Sect. 4 for climate simulations, (ii) how large the differences to the driving data are, and (iii) whether there are any trends in the accuracy of the results during the integration period, indicating spin-up effects or decoupling between the regional model and its large-scale forces at the lateral boundaries.

For the model evaluation, we concentrated on near-surface climate parameters whose regional representation and changes in future climate scenarios provide essential information on the expected impacts. We investigated near-surface temperature and precipitation, as in most other recent regional climate model inter-comparison projects.

We used the version of the multivariate clustering technique that is applicable to gridded data and compared the model results to the driving ECMWF re-analyses on the model grid. We utilized the prognostic variables temperature and specific humidity at the lowest model level above the surface and performed the cluster analysis for the mean values for the entire 10-year simulation period of these model variables. Figure 9 shows the resulting patterns of differences between the clusters as found for the ERA data and those attributed to the model results at the individual grid points, expressed by $R^{\mathrm{norm}}_{\mathrm{Mod,Ref}}$ as described in Sect. 4.

In parts of the model region, a good agreement between the simulation results and ERA data is evident with cluster overlaps between 85–100% that correspond to an error measure between 0 and 15%. Over the Atlantic region, however, wave-like structures of stronger differences are visible with nearly no overlaps between the clusters as identified for the two data sets. Also, large regions over southeastern Europe can be observed with clear discrepancies between model results and analyses. In the subsequent diagnosis described below for which we utilized the developed module to evaluate 4-dimensional model variables on levels above the surface, we were later able to identify the causes of these two dominant features.

It has to be stressed that the applied algorithm is quite sensitive especially for small deviations. An error measure of 100% means that the two clusters for the model results and the reference data set have no overlaps and they are therefore different for the grid point of concern. No statement can be derived, however, on the real distance between the two clusters with this version of the method so far. They can be neighbors or belong to totally different regions of the distribution function for all clusters.

As in Case Study I, we investigated the individual near-surface parameters with standard and advanced univariate statistics in the next step. We performed this diagnosis, however, for near-surface temperature and precipitation instead for their prognostic proxies that we utilized for the cluster analysis. This allows us to use more than one reference data set and to assess the differences between such reference data sets. In particular, we compared the model results not only to the ERA data, but also to a gridded observations data set that was generated by the Climate Research Unit (CRU) at the University of East Anglia (New et al., 2000). Because this reference data set is available for land areas only, we excluded oceanic regions from these investigations.

In Fig. 10, the spatial patterns for the near-surface 2-meter temperature (T2m) are shown. Although the figure shows comparable structures, no trustworthy conclusions on the quality of the model results can be drawn from a visual comparison between simulation results and reference data alone. Instead, such measures as the bias between these two arrays should be used as shown in Fig. 11. In this representation, remarkable differences become obvious, especially in the northern and eastern parts of the model region that can hardly be diagnosed from Fig. 10. This clearly indicates that a simple comparison of model results and reference data may lead to an overestimation of the model performance.

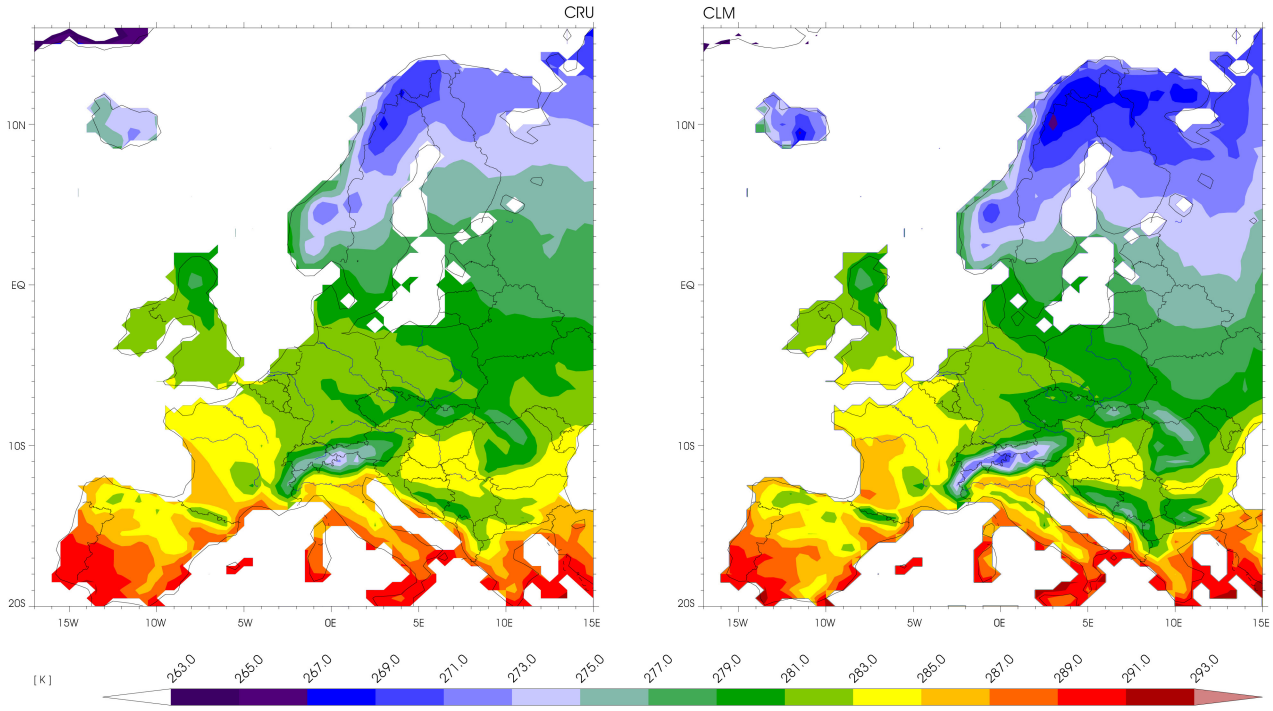

Fig. 10. Patterns of 2-m temperature over land, averaged for the 10-year simulation period 1979–1988. Left panel: gridded CRU data set, right panel: CLM model results (run0).

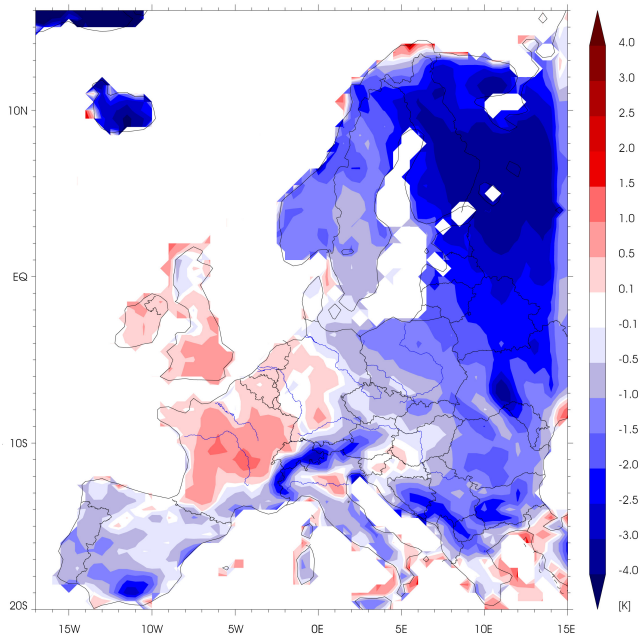

Fig. 11. Mean bias of 2-m temperature over land between CLM model results (run0) and gridded CRU data set for the 10-year simulation period 1979–1988.

Table 2. Key figures for two 10-year CLM regional climate simulation runs over Europe.
CLM run0 – basic run.
CLM run1 – run with modified initial and boundary data.
Acronyms are explained in the text.

	CLM run0	CLM run1
BIAS [K]	–1.05	–1.13
RMSE [K]	1.56	1.60
PACO [1]	0.98	0.99
RSV [1]	1.41	1.43
ROYA [1]	1.17	1.15
RTV [1]	1.41	1.38
TCO [1]	1.00	1.00

For an unambiguous quantitative assessment of the performance of the used model, we apply key figures that represent distance measures for various aspects of the quality of the simulation results. In addition to the evaluation of individual model runs, they are also applicable for an objective model intercomparison or for a comparison of different experiments with the same model. Complementary to difference plots, they are required to judge the success of model modifications.

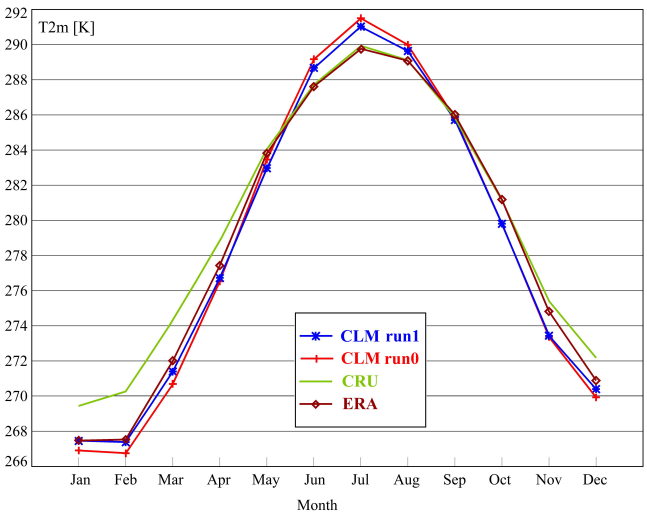

Fig. 12. Mean annual cycle of area-averaged 2m-temperature (T2m) for land grid points only.

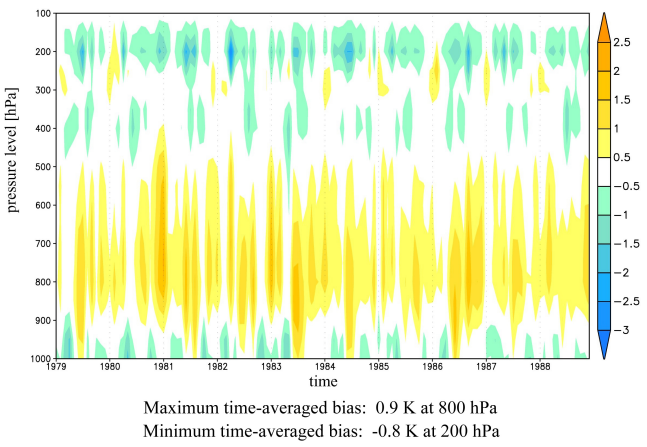

Fig. 13. Bias CLM run0–ERA of monthly mean, area averaged temperature [K] on pressure levels.

For the evaluation of the simulated 2-m temperature, we calculated such a set of key figures using the gridded CRU reference data set for all land grid points (Table 2).

The BIAS indicates that the modeled temperatures are on average 1.0 K below the observations for the whole model region, whereby error compensation may take place because of this averaging. The root mean square error (RMSE) as an area-averaged measure of the total differences, however, turns out to be similar to the BIAS and suggests that this sign-related error compensation is negligible. The pattern correlation coefficient (PACO) between the temporal averages of the annual means for the two data sets is 0.98 and proves that the temperature patterns compare well. The ratio of the spatial variances (RSV) of 1.41 hints at a stronger spatial variability in the model results. The last three quantities measure the capability of the model to reproduce the observed temporal variability. ROYA is the ratio of the yearly amplitudes, RTV

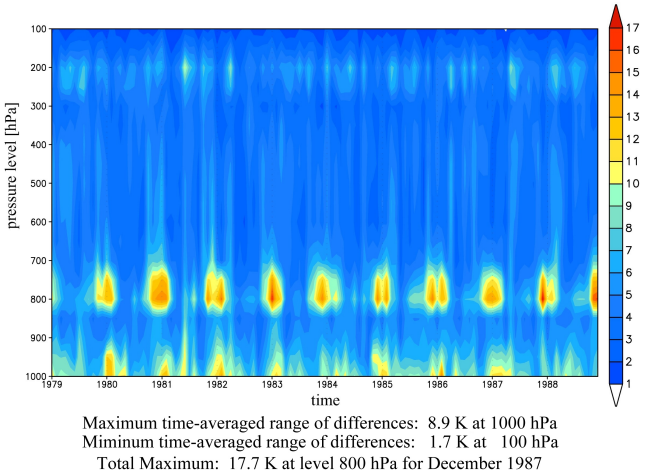

Fig. 14. Range of pattern differences CLM run0–ERA of monthly mean, averaged temperature [K] on pressure levels.

describes the ratio of the temporal variances in model results and reference data, and TCO is defined as the temporal correlation of the spatially averaged annual cycle between model and reference data. In Table 2, all used individual measures are listed.

Altogether, we found that the spatial and temporal variability in the model results is higher than the one in the CRU reference data set, but also that a very close correspondence between the annual cycles exists. The finding of stronger spatial structures in the model results may not necessarily be attributed to a model deficiency. Rather, this outcome could be linked to a smoothing effect in the reference data set during a thin-plate spline interpolation from the observation station sites to the used grid. The higher temporal variability in the model can be explained by a stronger annual cycle as visible in Fig. 12 (red line, CLM run0). It is obvious that the model slightly overestimates the temperatures during summer, but provides temperatures, which are too low in autumn and early winter. For springtime, the differences between the two reference data sets are larger than the ones between the model results and each of the reference data sets. Therefore, no clear conclusion can be derived for this season.

In the third step, we applied the developed evaluation module for 4-dimensional model variables aiming at identifying the sources for the differences between model results and reference data sets that were detected for one of the near-surface variables.

Figure 13 shows the temperature bias between CLM results and ERA data, now for the area mean for the entire model area. We excluded the boundary relaxation sponge zone from this analysis, because in this band, model results are relaxed towards the driving analysis data with decreasing weights and decreasing distance to the boundaries leading to an overestimation of the model performance there. It is visible that the dominating cold bias close to the surface during most of the simulation period is caused by too high temperatures at the levels above. This indicates an upward shift of

thermal energy in the model. The maximum is identified for the pressure level of 800 hPa, reaching up to 2.5 K there.

As mentioned earlier, differences at individual grid points may be underestimated due to averaging effects when using the bias alone. Figure 14 proves that this is true especially for vertical levels around 800 hPa. The total maximum of the deviations between model results and analysis data, expressed by the range of pattern differences, is still identified at 800 hPa, but the amplitude is much bigger now. Also, the maximum of the temporal averaged differences for the entire integration period is identified at 1000 hPa with this error measure instead at 800 hPa when using the bias.
We further investigated the spatial patterns on the pressure level of 800 hPa for December 1987 (not shown), the month with the maximum range of pattern differences for the entire simulation period, and found indications of an unrealistic representation of the analysis data over the Alps, although not for the model results in this region. A similar pattern was detected for the mean for the entire simulation period close to the surface, as shown in Fig. 15.

In the lower panel, a positive bias is obvious for most parts of Central Europe, showing a circular structure. In combination with diagnosis results for other atmospheric variables, we identified the source of these differences in the analysis data as an artificial modulation of the original data fields during preprocessing. After re-generating the initial and boundary data residing at another original spatial representation, the wave-like structures that were identified in the cluster analysis disappeared.

As in the case of too high spatial variability for the simulated 2-m temperature, this underlines that differences should not automatically be interpreted as model errors and that they may be linked to data transformation algorithms that are applied to the reference data.
The evaluation of the model re-run (CLM run1) using the corrected analysis data indicated, however, only small effects on the 2-m temperature. An overview of the individual key figures in Table 2 shows a slightly improved temporal variability in the model results as expressed by lower values of ROYA and RTV. This improvement is linked to a better representation of the annual cycle, especially during summer, as visible in Fig. 12 (blue line, CLM run1). BIAS and RMSE, however, reveal a somewhat increased systematic error on average for the entire simulation period.

There are hints that the remaining differences between model results and the utilized reference data may be linked to the representation of soil processes in the model. To get an impression on the sensitivity of CLM to the formulation of the soil process parameterizations and to the assignment of values to soil-related empirical constants, we started experiments in which we scanned a certain range of values for combinations of such empirical parameters using the multi-run simulation environment as described in the next section.

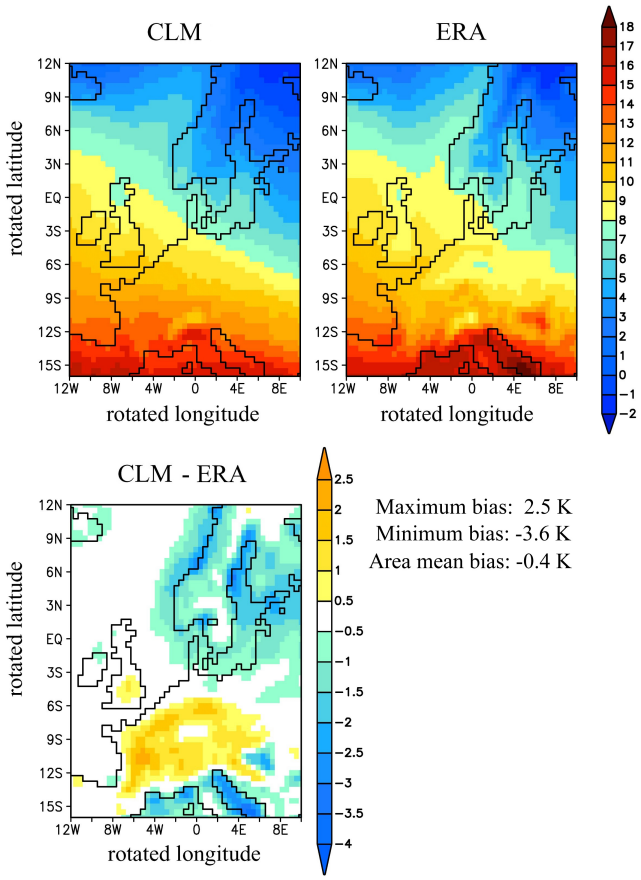

Fig. 15. 10-year mean of monthly mean temperature [°C] at 1000 hPa. Upper left panel: CLM results (run0), upper right panel: ERA data, lower panel: bias CLM–ERA [K]. Black lines: Model land-sea mask.

7 A multi-run simulation environment

Although multi-run experiments for long-term regional climate model simulations are actually still too time-consuming due to computational constraints, such experiments can be performed to investigate extremes and to analyze the behavior of the model for disturbed internal parameters on shorter time scales up to several months. A recently developed software environment (Flechsig et al., 2003) allows a model to be easily embedded into this environment and multiple model runs to be carried out in a coordinated manner. One of the supported experiment types is behavioral analysis, where a given range of meaningful values for combinations of model parameters in a user-defined way can be scanned. Further supported experiment types are local sensitivity analysis, where the sensitivity of a model to parameter variations in the vicinity of their default values can be investigated, and Monte-Carlo analysis with randomly changing parameter values following a certain distribution function. Especially for the latter, a set of experiment specific post-processing operators was developed to estimate among others lines of confidence, covariances, moments and further

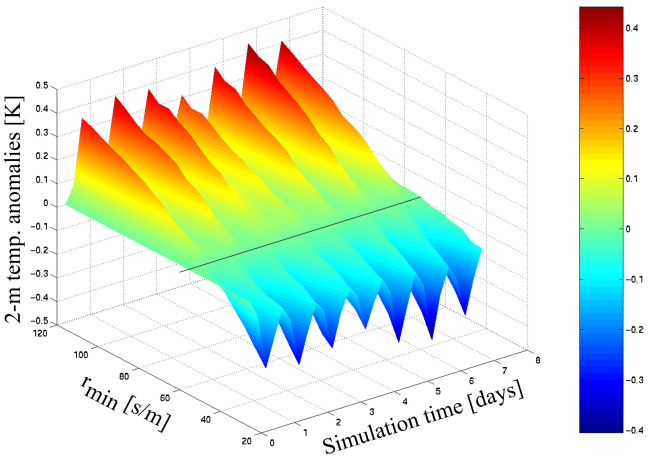

Fig. 16. CLM model results for a behavioral analysis. Anomalies of area means for T2m, depending on the minimum resistance r_{min} and the simulation time.

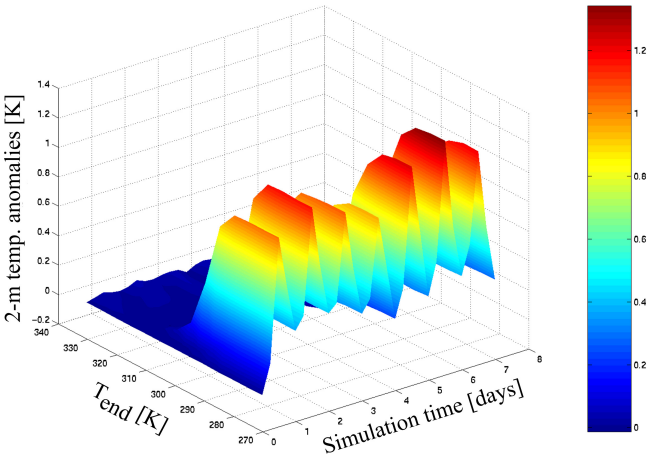

Fig. 17. CLM model results for a behavioral analysis. Anomalies of area means for T2m, depending on the upper temperature threshold for plant transpiration T_{end} and the simulation time.

attributes of the result distribution function as median, kurtosis or skewness, and quantiles, or to perform a regression. This simulation environment is still under development and integration of further operators is intended. Furthermore, users can define their own operators and operator pipelines easily.

To use this potential, we started to embed the regional climate model CLM into this environment. In a first step, we performed an analysis of the model's behavior depending on some of the internal parameters in the soil sub-model.

In this sub-model, evapotranspiration is computed by means of a simplified Biosphere-Atmosphere Transport Scheme (BATS) following the approach of Dickinson (1984). As the main idea, a resistance concept for the different processes influencing evapotranspiration is utilized with atmo-

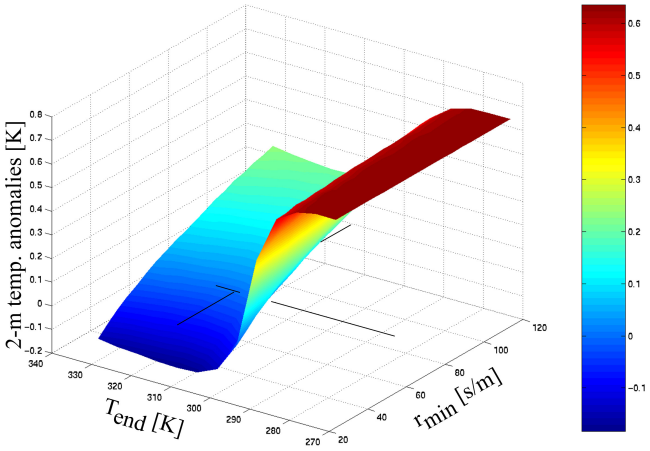

Fig. 18. CLM model results for a behavioral analysis. Temporal averaged anomalies of area means for T2m, depending on the upper temperature threshold for plant transpiration T_{end} and the minimum resistance r_{min}.

spheric, stomatal and leaf area-related components. For a behavioral analysis, we concentrated first on the stomatal resistance r_s which is computed by:

$$\frac{1}{r_s} = \frac{1}{r_{max}} + \left(\frac{1}{r_{min}} - \frac{1}{r_{max}}\right) \cdot F_{rad} \cdot F_{wat} \cdot F_{tem} \cdot F_{hum} \,. \quad (4)$$

where r_{min} and r_{max} represent scalable parameters that are set to default values of 90 s/m and 1000 s/m respective. $F_{<.>}$ are reduction functions for different factors (radiation, soil water, surface temperature and humidity) and cover a range between 1 (most favorable conditions) and 0 (total unfavorable transpiration conditions). The temperature-related reduction function permits plant transpiration in a range between T_0 (0 °C or 273.15 K) and an upper threshold value T_{end} according to equation (5) with the surface temperature T_b:

$$F_{tem} = \mathrm{Max}\left[0;\ \mathrm{Min}\left\{1;\ 4 \cdot \frac{(T_b - T_0) \cdot (T_{end} - T_b)}{(T_{end} - T_0)^2}\right\}\right] . \quad (5)$$

In a first experiment, we altered the default value for r_{min} between 60 s/m and 120 s/m. Figure 16 shows the influence of these modifications on T2m. In this representation, the temporal evolution of the anomalies of the area mean for T2m, related to the model run with the default parameter is displayed during the first week of a simulation starting in August 1995 over the Baltic sea region depending on r_{min}. As in the following figures, the color bar provides the range of values covered in the surface plot, and the black line denotes the default run (partially hidden). Model output is stored every 6 h and indicates a pronounced daily cycle in the anomalies. The major result is, however, a rather linear dependency of T2m on r_{min}. Increasing resistance values increase the near-surface temperatures due to a reduction of the latent heat flux and moisture transport by the plants.

A completely different behavior occurs for altered values of T_{end} in a second experiment. In this case, we varied T_{end} in the range between 273.15 K or 0°C and 333.15 K or 60°C. Figure 17 shows the behavior of T2m in the same representation as in Fig. 16. At higher than default values, nearly no sensitivity of the model results for T2m can be observed. In the lower parameter value range, however, a small region of strongly increasing temperatures due to reduced transport of heat by plant transpiration (not shown) can be observed. Furthermore, a distinct daily cycle is visible for the lower range of T_{end}.

The SimEnv environment allows not only the investigation of the response of a model to altered values of individual parameters, but also to combinations of them. We used this feature to perform an experiment in which we altered r_{min} and T_{end} combinatorially together in such a way that we altered T_{end} between 273.15 K and 333.15 K for each value of r_{min} between 60 s/m and 120 s/m. The result is shown in Fig. 18, where the anomalies of the area means for T2m in relation to the default model run, averaged for the first 7 days of the integration period are presented as a function of both parameters. As the most interesting result, a clear dominance of T_{end} for small values becomes visible with nearly no response of T2m to changed values of r_{min}, whereas for higher values of T_{end} the near-surface temperature increases with increasing values of r_{min}.

In the next steps, we aim at further experiments of this kind to scan a wider range of scalable parameters by local sensitivity analyses and to identify those parameters providing the major contributions to the variability in the model results. Further, we will focus on ensemble simulations for extreme periods with disturbed initial conditions. For these investigations, we use the possibility of SimEnv to perform Monte-Carlo experiments, where the initial conditions can be disturbed according to given distribution functions.

8 Conclusions

We presented two case studies as examples for the application of a regional climate model for the simulation of extremes and of long-term climatic characteristics.

In Case Study I, we investigated the risk of potential total yield loss for maize in the Northeast of Brazil during an El Niño-related drought. We performed a cluster analysis to evaluate the model results and defined the two generalized cost functions Q_1, which quantifies differences in the properties of the pairs of most similar clusters that were identified for model and reference data, and Q_2, which addresses structural differences. Using these measures, we found the quality of the simulation results in reproducing this extreme event in a range comparable to another regional model simulation, with a better representation of the spatial drought patterns in the region. The source of error in this simulation is an overestimation of precipitation primarily during February and March 1983. Excessive rainfall is simulated in coastal regions with the highest error for March 1983. This excess appears to be linked to an overestimation of vertical mass transport near South America.

In Case II, we simulated the climatic conditions between 1979 and 1988 over Europe using the same model. Accuracy was assessed using a cluster analysis of a combined set of parameters describing temperatures and moisture in the lowest model level as proxies for the diagnostic near-surface variables. Differences between model results and ERA reference data are most pronounced over the Atlantic region and over parts of southeastern Europe. These differences are partially linked to a spurious modulation of the original analysis fields during preprocessing. Spatial patterns of 2-m temperature were reasonably well simulated. There is a slightly higher spatial variability in the model results than in the CRU gridded observations. However, this appears to be linked to excessive smoothing in these reference data rather than being a model deficiency. The amplitude of the annual cycle of 2-m temperatures is generally too large. A model re-run using corrected initial and boundary conditions indicated only minor effects on the results for the near-surface temperature.

For both case studies, we applied a newly developed common evaluation approach. The algorithm contains a module for multidimensional pattern recognition for sets of parameters which are either suitable to describe an extreme, or which characterize combinations of model variables and their statistical properties simultaneously. The elaborated general quality measures could be proofed to be applicable to the presented case studies and gave sensitive quantitative evaluation results.

In a further module, we use an extensible basic set of univariate standard measures and new key figures. They allow to unambiguously quantify the model performance for individual variables in terms of such characteristics as differences and ratios between spatial and temporal means and variances, and the associated correlations. This way, an objective intercomparison of different models is possible, and the success of model modifications or changes in the experiment setup may be judged.

In a third module, we implemented techniques to investigate the question, of whether a model is able to represent the atmospheric physical processes correctly. For this, we evaluate 4-dimensional model variables in terms of the differences between their area means, the magnitude of their highest differences at individual grid points and their pattern correlation. The algorithm is open for extensions to assess further diagnosis goals.

Altogether, the developed evaluation algorithm gave reasonable and interpretable results and demonstrated its applicability to this kind of research questions. It is, however, as a result of its conceptual design not restricted to such investigations, and can also serve as a test bed with a standard set of quality measures for other regional climate models.

Furthermore, we have started to implement the regional climate model CLM into a multi-run environment that allows to perform sensitivity and behavioral analyses, and Monte-Carlo simulations. Early results for experiments in which we

have investigated the behavior of the model to variations in parameters for the soil sub-model are encouraging.

Our next activities will focus on the one hand on adding more user interactivity to that module of the evaluation package, which assesses the quality of 4-dimensional atmospheric variables. This shall enable to analyze all prognostic variables together at crucial points in space and time indicating the strongest simulation errors. On the other hand, we aim at implementing additional statistics for extremes and multi-run experiments, and to complete the integration of the regional climate model CLM into the ensemble simulation environment.

Acknowledgements. We are grateful to the Deutscher Wetterdienst for provision of the Local Model and support during its implementation and adaptation to climate simulations. We are also grateful to the ECMWF for providing ERA data, and to H. Österle for his support in the extraction and compilation of these. Furthermore, we thank the Climate Research Unit at the University of East Anglia for providing their gridded data set. Also, the two anonymous reviewers are acknowledged for their constructive comments and valuable hints to re-structure the paper and to place more emphasis on the applicative aspects of the investigations.

Edited by: T. Glade
Reviewed by: two referees

References

Böhm, U.: Eine Methode zur Validierung von Klimamodellen für die Klimawirkungsforschung hinsichtlich der Wiedergabe extremer Ereignisse, PhD thesis, Freie Universität Berlin, Department of Earth Sciences, 33–40, 1999.

Böhm, U., Hauffe, D., and Kücken, M.: The New Local Model (LM) of the Deutscher Wetterdienst, Part III: User Guide – The preparation of ECMWF re-analyses and the interpolation program era2lm for nested long-term model simulations. Potsdam-Institut für Klimafolgenforschung, Department Climate System, 2002.

Böhm, U., Gerstengarbe, F.-W., Hauffe, D., Kücken, M., Österle, H., and Werner, P.C.: Dynamic Regional Climate Modeling and Sensitivity Experiments for the Northeast of Brazil, in: Global Change and Regional Impacts, edited by Gaiser, T., Krol, M., Frischkorn, H., and de Araújo, J. C.: G. Springer, Berlin, Heidelberg, New York, 153–170, 2003.

Christensen, J. H., Machenhauer, B., Jones, R. G., Schär, C., Ruti, P. M., Castro, M., and Visconti, G.: Validation of present-day regional climate simulations over Europe, LAM simulations with observed boundary conditions, Clim. Dynam., Vol. 13, No. 7-8, 489–506, 1997.

Dickinson, R.: Modeling Evapotranspiration for Three-dimensional Global Models, in: Climate Processes and Climate Sensitivity. Geophysical Monograph 29, Maurice Ewing Vol. 5, American Geophysical Union, Washington, D.C., 58–72, 1984.

Doms, G. and Schättler, U.: The Nonhydrostatic Limited-Area Model LM (Local Model) of DWD, Part I: Scientific Documentation, Deutscher Wetterdienst, Business Unit Research and Development, 1999.

Flechsig, M., Böhm, U., Nocke, T., and Rachimow, C.: The Multi-Run Simulation Environment SimEnv., User's Guide for Version V1.03, www.pik-potsdam.de/topik/pikuliar/simenv/home/, 2003.

Forgy, E. W.: Cluster Analysis of multivariate data: efficiency versus interpretability of classifications. Biometrics 21, 768-769, 1965.

Frei, C., Christensen, J. H., Déqué, M., Jacob, D., Jones, R. G., and Vidale, P. L.: Daily Precipitation Statistics in Regional Climate Models: Evaluation and Intercomparison for the European Alps, J. Geo. Res.-A., Vol. 108, No. D3, 4127–4130, 2003.

Gerstengarbe, F.-W. and Werner, P. C.: A method for the statistical definition of extreme-value regions and their application to meteorological time series, Z. Meteorol., 39, (4), 224–226, 1989.

Gerstengarbe, F.-W. and Werner, P. C.: A Method to Estimate the Statistical Confidence of Cluster Separation, Theor. App. C., Vol. 57, No. 1–2, 103–110, 1997.

Gerstengarbe, F.-W., Werner, P. C., Fraedrich, K.: Applying non-hierarchical cluster analysis algorithms to climate classification: some problems and their solution, Theor. App. C., Vol. 61, 143–150, 1999.

Gibson, J. K., Kallberg, P., Uppala, S., Hernandez, A., Nomura, A., and Serrano, E.: ECMWF re-analysis project report series, 1. ERA description, European Centre for Medium-range Weather Forecasts (ECMWF), Reading, UK, 1997.

Havre, S., Hetzler, E., Whitney, P., and Nowell, L.: Theme River: Visualizing thematic changes in large document collections, ACM T. Graph., Vol. 8, No.1, 9–20, 2002.

IPCC: Climate Change 2001: Contribution of Working Group I to the third Assessment Report of the IPCC, Cambridge University Press, 583–638, 2001.

Jacob, D., Claussen, M., Majewski, D., Podzun, R., and Rockel, B.: REMO – a model for climate research and weather forecast, Research activities in Atmospheric and Oceanic Modelling, WMO/JCSU/IOC Report, 1995.

Jones, R. G., Murphy, J. M., and Noguer, M.: Simulation of Climate Change over Europe using a nested regional climate model, I: Assessment of control climate, including sensitivity to location of lateral boundaries, Q. J. R. Meteo., Vol. 121, 1413–1449, 1995.

Kücken, M., Gerstengarbe, F.-W. and Werner, P. C.: Cluster analysis results of regional climate model simulations in the PIDCAP period, Boreal Env. Res., Vol. 7, No 3, 219–223, 2002.

Kücken, M. and Hauffe, D.: The Nonhydrostatic Limited Area Model LM (Local Model) of DWD with PIK extensions, Part III: Extensions User Guide, Potsdam-Institut für Klimafolgenforschung, 2002.

Machenhauer, B., Windelband, M., Botzet, M., Jones, R. G., and Déqué, M.: Validation of present-day regional climate simulations over Europe: Nested LAM and Variable Resolution Global Model Simulations wit Observed or Mixed Layer Ocean Boundary Conditions, Report No. 191, Max-Planck-Institut für Meteorologie, Hamburg, 9–23, 1996.

Machenhauer, B., Windelband, M., Botzet, M., Christensen, J. H., Déqué, M., Jones, R. G., Ruti, P. M., and Visconti, G.: Validation and Analysis of Regional Present-Day Climate and Climate Change Simulations over Europe, Report No. 275, Max-Planck-Institut für Meteorologie, 26–54, 1998.

New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability, part II, Development of 1901-96 monthly grids of terrestrial surface climate, J. Climate, Vol. 13, 2217–2238, 2000.

Nocke, T., Böhm, U., Schumann, H., and Flechsig, M.: Information Visualization Supporting Modeling and Evaluation Tasks For Climate Models, in: Proceedings of the 2003 Winter Simulation Conference, December 7–10, New Orleans, edited by Chick, S., Sanchez, P. J. S., Ferin, D., and Morrice, D.J., 2003.

Pettitt, A. N.: A Non-parametric Approach to the Change-Point Problem. Applied Statistics 28, No.2, 126–135, 1979.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T.: Numerical Recipes, The Art of Scientific Computing, Cambridge University Press, Chapter 13, 1986.

Roeckner, E., Arpe, K., Bentsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schleese, U., and Schulzweida, U.: The atmospheric general circulation model ECHAM-4, Max-Planck-Institut für Meteorology, Report No. 218, 1996.

Steinhausen, D. und Langer, K.: Clusteranalyse: Eine Einführung in Methoden und Verfahren der automatischen Klassifikation, Walter-deGruyter-Verlag, 100–138, 1977.

Takle, E. S., Gutowski Jr., W. J., Arritt, R. W., Pan, Z., Anderson, C. J., Silva, R., Caya, D., Chen, S.-C., Christensen, J. H., Hong, S.-Y., Juang, H.-M. H., Katzfey, J. J., Lapenta, W. M., Laprise, R., Lopez, P., McGregor, J., and Roads, J. O.: Project to Intercompare Regional Climate Simulations (PIRCS): Description and initial results, J. Geo. Res.-A., Vol. 104, No. D16, 19 443–19 462, 1999.

Taubenheim, J.: Statistische Auswertung geophysikalischer und meteorologischer Daten, Akademische Verlagsgesellschaft Geest & Porting, Leipzig, 1969.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geo. Res.-A., Vol. 106, No. D7, 7183–7192, 2001.